

ARBEITSBERICHTE DES INSTITUTS FÜR MATHEMATISCHE MASCHINEN UND DATENVERARBEITUNG

Prof. Dr. Wolfgang Händler

FRIEDRICH ALEXANDER UNIVERSITÄT ERLANGEN NÜRNBERG

Optimale Vorhersage der Ausgabe stochastischer Automaten über lange Zeiten

Alfred Schmitt

**Von der Technischen Fakultät der Friedrich-Alexander-Universität
in Erlangen-Nürnberg angenommene Habilitationsschrift**

BAND 3 · NUMMER 6 · ERLANGEN · FEBRUAR 1971

ARBEITSBERICHTE

DES INSTITUTE FÜR MATHEMATISCHE
MASCHINEN UND DATENVERARBEITUNG

Hrsg. v. Wolfgang Müller

HEFT 2. QUARTAL 1968 (JANUAR) 10. JAHRE

Optimale Vervollständigung der Aussagen
mathematischer Aussagen über lange Folgen

Alfred Schmitt

Von der Fakultät für Mathematik der Universität Bonn
an der Mathematischen Arbeitsgruppe der Universität Bonn
am 1. April 1968 in Bonn

105-116 161



Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41548403_0088

ARBEITSBERICHTE DES INSTITUTS FÜR MATHEMATISCHE MASCHINEN UND DATENVERARBEITUNG

Prof. Dr. Wolfgang Händler

FRIEDRICH ALEXANDER UNIVERSITÄT ERLANGEN NÜRNBERG

Optimale Vorhersage der Ausgabe stochastischer Automaten über lange Zeiten

Alfred Schmitt

Von der Technischen Fakultät der Friedrich-Alexander-Universität
in Erlangen-Nürnberg angenommene Habilitationsschrift

BAND 3 · NUMMER 6 · ERLANGEN · FEBRUAR 1971

ARBEITSBERICHTE
DES INSTITUTS FÜR MATHEMATISCHE
MASCHINEN UND DATENVERARBEITUNG

Prof. Dr. Wolfgang Sauer

TECHNISCHE UNIVERSITÄT ERLANGEN-NÜRNBERG

Optimale Vorhänge der Ausgabe
stochastischer Automaten über lange Zeiten



Druck HOGL Erlangen, Hauptstr. 109

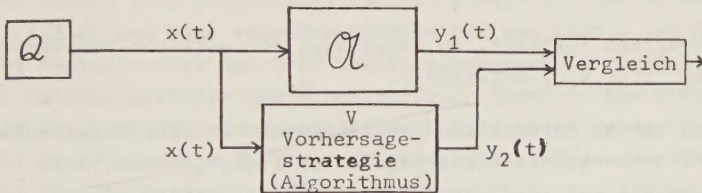
I n h a l t

	Seite
Einleitung	3
1. Stochastische Moore-Automaten, Allgemeines	
1.1 Einführung	8
1.2 Serien- und Parallelschaltung von Automaten	12
1.3 Matrizenkalkül für Automaten und Quellen	15
2. Vorhersage der Ausgabe eines Automaten	
2.1 Die Metrik Δ	19
2.2 Die Vorhersageanordnung	19
3. Optimale Vorhersage eines Automaten	
3.1 Die optimale Vorhersagestrategie V_{opt} für einen Automaten	27
3.2 Der Raum der Dichten	30
3.3 Q - α -Dichten und der Taktoperator	37
3.4 Die Existenz des mittleren Vorhersagefehlers bei optimaler Vorhersage	44
4. Vorhersage von Automaten mit Automaten	
4.1 Multiplex-Technik bei Automaten	50
4.2 Die Ausgabeverteilungen der Multiplex-Automaten	54
4.3 Vorhersage eines Automaten mit Multiplex-Automaten	61
Literatur	66
Begriffsregister	67
Wichtige Formelzeichen	69

Einleitung

In dieser Arbeit wird das Problem behandelt, die Ausgabe eines stochastischen Automaten bei Kenntnis der Eingabe möglichst gut vorherzusagen. Dabei werfen wir im wesentlichen die folgende Frage auf: Gegeben sei ein beliebiger stochastischer Automat, der Signale $x \in X$ als Eingaben entgegennimmt und jedes Signal in Abhängigkeit vom aktuellen inneren Zustand mit einem Ausgabesignal $y \in Y$ beantwortet. Wie gut kann bei Beobachtung einer langen Folge von Eingabesignalen die Ausgabe des Automaten vorhergesagt werden?

Um eine genauere Formulierung dieser Frage zu erreichen, betrachten wir die folgende Versuchsanordnung:



Die Quelle Q sei eine endliche Markow-Quelle, sie liefert in jedem diskreten Zeittakt $t = 1, 2, 3, \dots$ ein Signal $x(t) \in X$. Das Signal $x(t)$ läuft (zum Zeitpunkt t) als Eingabe in den Automaten $\mathcal{A}$ und erzeugt eine Ausgabereaktion $y_1(t)$. Die Vorhersagestrategie V liefert (in Abhängigkeit von $x(1), x(2), \dots, x(t)$) ein Signal $y_2(t)$. $y_2(t)$ soll nach Möglichkeit so gewählt werden, daß $y_2(t) = y_1(t)$. Da der Automat stochastisch, also nicht-deterministisch ist, läßt sich die Gleichheit $y_2(t) = y_1(t)$ nur mit gewissen Wahrscheinlichkeiten erzielen. Wenn wir einmal annehmen, daß die gesamte Versuchsanordnung immer wieder im Urzustand, also unter den gleichen Anfangsbedingungen, gestartet wird, so ist klar, daß dabei immer wieder verschiedene Abläufe

erwartet werden müssen, denn weder Q noch $\mathcal{A}$ ist deterministisch. Die Funktionsweise von Quelle und Automat ist nur durch Wahrscheinlichkeiten festgelegt. Es existiert jedoch

$f(\mathcal{A}, V, Q, t) :=$ Erwartungswert (=Wahrscheinlichkeit) dafür, daß zum Zeitpunkt t $y_1(t) = y_2(t)$ gilt, unter der Voraussetzung, daß die Versuchsanordnung im Urzustand gestartet wird.

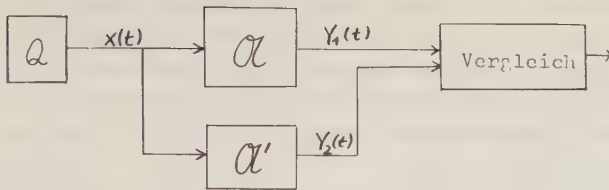
Dieser Erwartungswert gibt also Auskunft darüber, wie gut der Automat $\mathcal{A}$ bei Speisung durch Q zum Zeitpunkt t vorhergesagt wird. Für einen einzelnen Zeitpunkt ist die Betrachtung des momentanen Vorhersagefehlers nicht sonderlich interessant. Wir streben vielmehr eine zeitunabhängige Charakterisierung der Güte einer Vorhersage an. Dies gelingt durch die Definition

$$\tilde{F}(\mathcal{A}, V, Q) := \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T f(\mathcal{A}, V, Q, t).$$

Falls Konvergenz vorliegt, so sagen wir, der mittlere Vorhersagefehler $\tilde{F}(\mathcal{A}, V, Q)$ existiert.

Damit sind wir am eigentlichen Ausgangspunkt unserer Untersuchung angelangt. Der mittlere Vorhersagefehler $\tilde{F}(\mathcal{A}, V, Q)$ gibt also an, wie gut die Ausgaben des Automaten $\mathcal{A}$ bei Speisung durch Q von der Vorhersagestrategie V vorhergesagt werden können, wobei die betrachteten Zeiträume nicht begrenzt sind. Wir lassen sogar Strategien V zu, die ebenfalls stochastisch sind, d. h. die das vorhergesagte Signal $y_2(t)$ nur mit bestimmten Wahrscheinlichkeiten festlegen.

Wie bereits angedeutet, existiert $\tilde{F}(\mathcal{A}, V, Q)$ - mangels Konvergenz - nicht für jede beliebig gewählte Strategie V . In Abschnitt 2 können wir jedoch bereits beweisen, daß $\tilde{F}(\mathcal{A}, V, Q)$ stets existiert, wenn sich V durch einen stochastischen Automaten realisieren läßt, d. h. wenn wir eine Anordnung der der Gestalt



vorliegen haben.

Falls $\tilde{F}(\mathcal{A}, V, Q)$ existiert, so gilt $0 \leq \tilde{F}(\mathcal{A}, V, Q) \leq 1$. Die Strategie V ist offenbar als umso besser einzuschätzen, je kleiner $\tilde{F}(\mathcal{A}, V, Q)$ ist. Die Größe

$$\tilde{F}_{\min}(\mathcal{A}, Q) := \inf_V \tilde{F}(\mathcal{A}, V, Q)$$

stellt eine untere Schranke dar, die durch keine Vorhersagestrategie V unterboten werden kann. Das Infimum ist dabei nur über solche V zu erstrecken, für die $\tilde{F}(\mathcal{A}, V, Q)$ auch existiert. $\tilde{F}_{\min}(\mathcal{A}, Q)$ ist eine bedeutsame Größe für das Paar Q und $\mathcal{A}$. Sie gibt Auskunft über den Grad der Vorhersagbarkeit und damit auch der Determiniertheit von $\mathcal{A}$ bei Speisung durch Q . Die Abhängigkeit von Q könnte nun ebenfalls noch eliminiert werden durch Wahl einer Standard-Quelle. Auf diese Weise würde man eine Maßzahl für die Determiniertheit von $\mathcal{A}$ erhalten. Diese Überlegungen verfolgen wir jedoch nicht weiter. Falls $\mathcal{A}$ ein deterministischer endlicher Automat ist, so hat das Vorhersageproblem eine triviale Lösung. Diesen Fall werden wir nicht gesondert diskutieren.

In Abschnitt 3 beweisen wir, daß zu jedem Automaten $\mathcal{A}$ eine optimale Vorhersagestrategie V_{opt} existiert mit

$$\tilde{F}(\mathcal{A}, V_{\text{opt}}, Q) = \tilde{F}_{\min}(\mathcal{A}, Q).$$

Die Hauptschwierigkeit des Beweises liegt darin, die Existenz von $\tilde{F}(\mathcal{A}, V_{\text{opt}}, Q)$ zu beweisen, denn V_{opt} ist im allgemeinen nicht mit einem endlichen Automaten realisierbar.

In Abschnitt 4 werfen wir daher die Frage auf, ob sich nicht auch bereits mit endlichen Mitteln noch eine hinreichend gute Vorhersage erzielen läßt, obwohl wir den Vorhersagezeit-

raum ja nicht begrenzen. Diese Frage ist auch aus technischer Sicht von Interesse, denn zur Lösung einer Vorhersageaufgabe werden im Normalfall eben nur stochastische Automaten zur Verfügung stehen. Unter Zuhilfenahme einer speziellen Multiplex-Technik für stochastische Automaten gelingt uns der Beweis für den folgenden Satz:

Zu jedem Automaten $\mathcal{A}$ mit speisender Quelle Q und jedem $\varepsilon > 0$ existiert ein endlicher stochastischer Automat $\mathcal{A}'$ mit

$$\tilde{F}(\mathcal{A}, v_{\mathcal{A}'}, Q) - \tilde{F}_{\min}(\mathcal{A}, Q) < \varepsilon ,$$

dabei ist $v_{\mathcal{A}'}$ die mit $\mathcal{A}'$ realisierbare Vorhersagestrategie.

Dies besagt also, daß man stochastische Automaten mit stochastischen Automaten fast beliebig gut - im Rahmen des überhaupt Möglichen - vorhersagen kann. Der einzige mittlere Vorhersagefehler, den man so evtl. nicht erreichen kann, ist $\tilde{F}_{\min}(\mathcal{A}, Q)$.

Die Multiplex-Technik liefert dabei einen Hinweis darauf, wie aus einem gegebenen unzuverlässigen stochastischen Automaten zuverlässigere Automaten gebaut werden können, bei denen die Störungen, also die weniger wahrscheinlichen Ereignisse, immer stärker zugunsten der wahrscheinlichsten Ereignisse zurückgedrängt werden. Damit ergeben sich z. B. Berührungspunkte zu von NEUMANNs klassischer Arbeit "Probabilistic Logics and the Synthesis of Reliable Organisms From Unreliable Components" [7].

Ein stochastischer Automat kann nach einem Satz von NIEH und CARLYLE ([5], vgl. auch Abschnitt 4.3) auch dargestellt werden als ein deterministischer Automat mit einer eingebauten störenden Rauschquelle. Insofern sind unsere Untersuchungen auch relevant zu Problemen der Gehirnstromforschung, der gestörten Nachrichtenübertragung (Informationstheorie) und der Informationsverarbeitung in endlichen Systemen.

Es sei weiter angemerkt, daß wir bei allen unseren Beweisen

dem Automatenmodell und der Quelle - außer der Endlichkeit - keine Einschränkungen auferlegt haben. Eigenschaften wie z. B. die Ergodizität vereinfachen in manchen Fällen die Beweise wesentlich, worauf wir aber bewußt verzichtet haben, um allgemeingültige Resultate zu erzielen. Wir haben ebenfalls darauf verzichtet, die Probleme des Sampling zu diskutieren, also die Frage, wie aus einem einmaligen Lauf der Vorhersageanordnung unter Registrierung der Fehlerhäufigkeit auf den tatsächlichen mittleren Vorhersagefehler der Anordnung geschlossen werden kann. Im allgemeinen läßt sich so der mittlere Vorhersagefehler nämlich nicht ermitteln.

Ähnliche Untersuchungen wie in dieser Arbeit wurden von BOOTH [1] und WOLF [8] durchgeführt. Den Arbeiten von BOOTH und WOLF liegen jedoch andere Vorhersageanordnungen und -ziele zugrunde.

1. Stochastische Moore-Automaten, Allgemeines

1.1 Einführung

In diesem Abschnitt wollen wir die Grundbegriffe aus der Theorie der stochastischen Automaten zusammenstellen und gleichzeitig unsere Terminologie einführen.

Ein Alphabet X ist stets eine endliche, nichtleere Menge. Wenn X ein Alphabet ist, so bezeichnen wir mit X^* die Menge aller endlich langen Worte, die man aus Elementen von X bilden kann. Wenn $\bar{x} \in X^*$, so hat $\bar{x}$ also die Gestalt

$$\bar{x} = x^1 x^2 \dots x^s,$$

und für s gilt $s \geq 0$. Dabei bezeichnet x^i formal das i -te Symbol des Wortes oder der Kette $\bar{x}$. Wir bezeichnen

$$\ell(\bar{x}) = \ell(x^1 x^2 \dots x^s) := s$$

als die Länge von $\bar{x}$. Das sogenannte leere Wort, welches die Länge 0 hat, bezeichnen wir mit Λ .

Wenn $\bar{x}$ und $\bar{x}'$ Worte sind, so bezeichne $\bar{x}\bar{x}'$ dasjenige Wort, das entsteht, wenn man $\bar{x}$ und $\bar{x}'$ aneinanderkettet (konkateniert). Mit der Operation der Aneinanderkettung, die assoziativ ist, wird X^* zu einer Halbgruppe, die auch als die "freie, von X erzeugte Halbgruppe" bezeichnet wird. Λ ist Einselement dieser Halbgruppe.

Gelegentlich benutzen wir die Bezeichnung X^k für die Menge

$$X^k := \{ \bar{x} \mid \bar{x} \in X^* \wedge \ell(\bar{x}) = k \}$$

aller Worte oder Ketten genau der Länge k .

Sei M eine Menge. Eine Abbildung $f: M \rightarrow [0,1]$ von M in das abgeschlossene reelle Intervall zwischen 0 und 1 heißt ein diskretes Wahrscheinlichkeitsmaß (W-Maß) über M , wenn

$$\sum_{\substack{m \in M \\ f(m) > 0}} f(m) = 1 \quad .$$

Die Menge der diskreten W-Maße über M bezeichnen wir mit $P(M)$. Ein Element $f \in P(M)$ ist also eine Abbildung $f: M \rightarrow [0,1]$ mit den oben gesetzten Einschränkungen.

Seien M und M' Mengen. Eine Abbildung

$$w: (M) \times (M') \rightarrow [0,1]$$

heißt bedingte diskrete Wahrscheinlichkeitsfunktion, wenn für alle $m' \in M'$ gilt

$$\sum_{m \in M} w(m|m') = 1.$$

$$w(m|m') > 0$$

$w(m|m')$ heißt die Wahrscheinlichkeit von m unter der Bedingung m' . Eine bedingte diskrete Wahrscheinlichkeitsfunktion nennen wir vereinfachend bedingte Wahrscheinlichkeit.

Definition 1.1

Ein initialer stochastischer Moore-Automat mit determinierter Markierung (im folgenden einfach Automat genannt) ist ein 6-Tupel

$$\mathcal{A} = (X, Y, Z, w, \mu, \pi_0),$$

in dem folgende Bestimmungsstücke aufgezählt sind:

- (1) X , ein Eingabealphabet (Eingabesignale),
- (2) Y , ein Ausgabealphabet (Ausgabesignale),
- (3) Z , eine endliche, nichtleere Menge von inneren Zuständen,
- (4) w , eine bedingte Wahrscheinlichkeitsfunktion der Gestalt $w: (Z) \times (Z \times X) \rightarrow [0,1]$,
- (5) μ , eine Ausgabeabbildung $\mu: Z \rightarrow Y$ (Markierungsfunktion) und
- (6) π_0 , ein (initiales) diskretes W-Maß über Z (initiale Zustandsverteilung).

Ein stochastischer Automat nimmt Eingabesignale entgegen, verändert dabei seinen inneren Zustand und gibt eine Ausgabe ab. Er arbeitet im allgemeinen nichtdeterministisch, jedoch ist mit w eine Angabe darüber gemacht, nach welchen Wahrscheinlichkeiten der Folgezustand sich beim Zustandsübergang einstellt. Zur Ver-

einfachung der Verhältnisse nehmen wir an, daß der Automat nur zu diskreten Zeitpunkten 1, 2, 3, ... eine Eingabe annimmt und gleichzeitig eine Ausgabe erzeugt. Wir setzen ferner voraus, daß sich der Automat zu jedem beliebigen Zeitpunkt in genau einem seiner Zustände $z \in Z$ befindet. Falls wir diesen Zustand nicht sicher, sondern nur mit gewissen Wahrscheinlichkeiten kennen, so benutzen wir Zustandsverteilungen, also diskrete W-Maße zur Beschreibung des Zustandes.

Definition 1.2

Die Wahrscheinlichkeit dafür, daß der Automat $\mathcal{A}$ (Def. 1.1) vom Zustand $z \in Z$ bei Eingabe des Signals $x \in X$ in den Zustand $z' \in Z$ übergeht, sei gleich $w(z'|z, x)$.

Die Wahrscheinlichkeit $w(y|z, x)$ dafür, daß $\mathcal{A}$ im Zustand $z \in Z$ bei Eingabe von $x \in X$ die Ausgabe $y \in Y$ liefert, sei festgelegt gemäß

$$w(y|z, x) := \sum_{z' \in Z \wedge \mu(z') = y} w(z'|z, x) .$$

Aus Definition 1.2 lassen sich mit einfachen wahrscheinlichkeitstheoretischen Rechenregeln auch Wahrscheinlichkeiten für längere Folgen von Eingabesignalen bestimmen:

Die Wahrscheinlichkeit $w(z^1 z^2 \dots z^s | z^0, x^1 x^2 \dots x^s)$ dafür, daß im Anfangszustand z^0 bei Eingabe von $x^1 x^2 \dots x^s$ (Kette von Eingabesignalen) genau die Zustandsfolge $z^1 z^2 \dots z^s$ durchlaufen wird, ist

$$w(z^1 z^2 \dots z^s | z^0, x^1 x^2 \dots x^s) = w(z^1 | z^0, x^1) \cdot w(z^2 | z^1, x^2) \dots w(z^s | z^{s-1}, x^s) .$$

Die Wahrscheinlichkeit $w(\bar{y} | z^0, \bar{x})$ dafür, daß der Automat im Anfangszustand z^0 bei Eingabe der Folge $\bar{x} = x^1 x^2 \dots x^s \in X^s$ die Ausgabefolge $\bar{y} = y^1 y^2 \dots y^s \in Y^s$ liefert, ist

$$w(\bar{y} | z^0, \bar{x}) = \sum_{\substack{z^1 z^2 \dots z^s \in Z^s \\ \wedge \mu(z^i) = y^i \ (i=1,2,\dots,s)}} w(z^1 z^2 \dots z^s | z^0, \bar{x}) .$$

Die Wahrscheinlichkeit $w(\bar{y}|\pi, \bar{x})$ dafür, daß der Automat in der Zustandsverteilung π bei Eingabe von $\bar{x} \in X^S$ die Ausgabe $\bar{y} \in Y^S$ liefert, ist

$$w(\bar{y}|\pi, \bar{x}) = \sum_{z \in Z} \pi(z) \cdot w(\bar{y}|z, \bar{x}) .$$

Wir setzen vereinfachend $w(\bar{y}|\bar{x}) := w(\bar{y}|\pi_0, \bar{x})$, d. h. die Angabe der initialen Zustandsverteilung des Automaten kann entfallen. Die zuletzt genannte bedingte Wahrscheinlichkeit $w(\bar{y}|\bar{x})$ hat bei unseren weiteren Überlegungen eine zentrale Bedeutung, denn sie bestimmt im wesentlichen das "Verhalten" des stochastischen Automaten, wenn der Anfangszustand gemäß der initialen Zustandsverteilung eingestellt ist. $w(\bar{y}|\bar{x})$ kann man mit einer gewissen Berechtigung als die "Übertragungsfunktion" des Automaten bezeichnen, wobei in dieser Ausdrucksweise ein möglicher Brückenschlag zur Systemtheorie angedeutet wird. Automaten mit der gleichen Übertragungsfunktion nennen wir gelegentlich auch äquivalent.

Oben wurde für mehrere, an sich verschiedene, bedingte Wahrscheinlichkeitsfunktionen die Bezeichnung w benutzt. Wir glauben, daß die eingesetzten Argumente in jedem vorkommenden Einzelfall eine eindeutige Bestimmung der gemeinten Wahrscheinlichkeit gestattet, wodurch sich die Einführung unterschiedlicher Bezeichnungsweisen erübrigt. Streng genommen müßten wir auch an dem w stets den Index $\mathcal{A}$ mitschleppen, was sich im allgemeinen ebenfalls erübrigt, da kaum Verwechslungen zu befürchten sind.

Das von uns benutzte Modell eines stochastischen Automaten ist das Moore-Modell, d. h. die Ausgabe des Automaten hängt nur von dem nach der Eingabe des einzelnen Signals erreichten Zustand ab (Abbildung μ). Es ist jedoch bekannt (vgl. STARKE [6]), daß man zu jedem allgemeinen initialen stochastischen Automaten $\mathcal{A}$ (in dieser Arbeit nicht definiert) mit einer Übertragungsfunktion $w_{\mathcal{A}}(\bar{y}|\bar{x})$ einen Automaten $\mathcal{A}'$ gemäß unserer Definition finden kann mit $w_{\mathcal{A}'}(\bar{y}|\bar{x}) = w_{\mathcal{A}}(\bar{y}|\bar{x})$ für alle $\bar{x} \in X^*$ und $\bar{y} \in Y^*$.

Außerdem weist man leicht nach, daß die deterministische Markierung $\mu: Z \longrightarrow Y$ durch eine allgemeinere nichtdeterministische Markierung $\mu: Z \longrightarrow P(Y)$ ersetzt werden kann, ohne daß dadurch

ein leistungsfähigeres Automatenmodell entsteht. Es gilt auch hier die Aussage: Jede Übertragungsfunktion, die durch einen Automaten mit nichtdeterministischer Markierung realisiert werden kann, ist auch in einem passend gewählten Automaten mit deterministischer Markierung realisierbar. Überdies ist ein Konstruktionsverfahren bekannt, mit dem aus einem Automaten mit $|Z|$ Zuständen ein äquivalenter und deterministisch markierter Automat mit im wesentlichen $|Z| \cdot |Y|$ Zuständen gewonnen werden kann.

1.2 Serien- und Parallelschaltung von Automaten

Definition 1.3

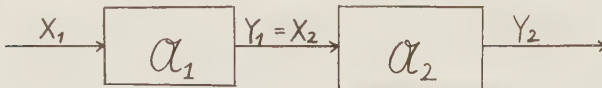
Seien $\mathcal{A}_i = (X_i, Y_i, Z_i, w_i, \mu_i, \pi_i)$ ($i = 1, 2$) zwei Automaten mit $Y_1 = X_2$. Dann ist das Serienprodukt $\mathcal{A}_1 \rightarrow \mathcal{A}_2$ der Automaten $\mathcal{A}_1$ vor $\mathcal{A}_2$ definiert als der Automat

$$\mathcal{A}_1 \rightarrow \mathcal{A}_2 = (X_1, Y_2, Z, w, \mu, \pi),$$

wobei

- (1) $Z := Z_1 \times Z_2$,
- (2) $w((z'_1, z'_2) | (z_1, z_2), x) := w_1(z'_1 | z_1, x) \cdot w_2(z'_2 | z_2, \mu_1(z'_1))$
für alle $z_1, z'_1 \in Z_1, z_2, z'_2 \in Z_2$ und $x \in X_1$,
- (3) $\mu(z_1, z_2) := \mu_2(z_2)$ für alle $z_1 \in Z_1, z_2 \in Z_2$ sowie
- (4) $\pi(z_1, z_2) := \pi_1(z_1) \cdot \pi_2(z_2)$ für alle $z_1 \in Z_1, z_2 \in Z_2$.

Die Serienschaltung der beiden Automaten $\mathcal{A}_1$ und $\mathcal{A}_2$ kann man strukturmäßig durch das Diagramm



veranschaulichen. Wenn $x \in X_1$ als Eingabe in das Produkt gegeben wird, so wirkt sie zunächst auf den Automaten $\mathcal{A}_1$, dieser geht in einen neuen Zustand über und sendet die entsprechende Ausgabe als Eingabe in $\mathcal{A}_2$. $\mathcal{A}_2$ geht nun seinerseits in einen neuen Zustand über und erzeugt dabei das Ausgabesignal, das durch die Eingabe von $x \in X_1$ in das Produkt bewirkt wird.

Für die Übertragungsfunktion w von $\alpha_1 \rightarrow \alpha_2$ gilt

$$w(\bar{y}|\bar{x}) = \sum_{\bar{y}' \in Y_1^* = X_2^*} w_1(\bar{y}'|\bar{x}) \cdot w_2(\bar{y}|\bar{y}') ,$$

wobei $\bar{x} \in X_1^*$, $\bar{y} \in Y_2^*$.

Die Übertragungsfunktion des Produktes läßt sich also ganz einfach aus den Übertragungsfunktionen der beteiligten Automaten gewinnen. Der Beweis dieser Aussage bereitet nur geringe Schwierigkeiten und wird hier übergangen.

Definition 1.4

Die Parallelschaltung $\alpha_1 \parallel \alpha_2$ der beiden Automaten

$$\alpha_i = (X_i, Y_i, Z_i, w_i, \mu_i, \pi_i) \quad (i = 1, 2)$$

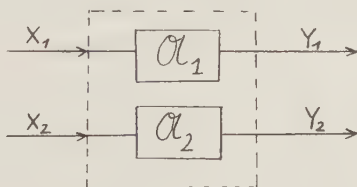
ist der Automat

$$\alpha_1 \parallel \alpha_2 := (X_1 \times X_2, Y_1 \times Y_2, Z_1 \times Z_2, w, \mu, \pi),$$

mit

- (1) $w((z'_1, z'_2) | (z_1, z_2), (x_1, x_2)) := w_1(z'_1 | z_1, x_1) \cdot w_2(z'_2 | z_2, x_2)$
für alle $z_1, z'_1 \in Z_1$, $z_2, z'_2 \in Z_2$, $x_1 \in X_1$ und $x_2 \in X_2$,
- (2) $\mu(z_1, z_2) := (\mu_1(z_1), \mu_2(z_2))$ für alle $z_1 \in Z_1$, $z_2 \in Z_2$ und
- (3) $\pi(z_1, z_2) := \pi_1(z_1) \cdot \pi_2(z_2)$ für alle $z_1 \in Z_1$ und $z_2 \in Z_2$.

Eine Veranschaulichung der Parallelschaltung ist durch das Diagramm



möglich. Beide Automaten werden als ein Ganzes aufgefaßt und parallel betrieben, es erfolgt aber keine gegenseitige Beeinflussung.

Es ist unmittelbar einzusehen, daß für die Übertragungsfunktion w der Parallelschaltung die Beziehung

$$w((y_1^1, y_2^1)(y_1^2, y_2^2) \dots (y_1^s, y_2^s) | (x_1^1, x_2^1)(x_1^2, x_2^2) \dots (x_1^s, x_2^s)) = \\ w_1(y_1^1 y_1^2 \dots y_1^s | x_1^1 x_1^2 \dots x_1^s) \cdot w_2(y_2^1 y_2^2 \dots y_2^s | x_2^1 x_2^2 \dots x_2^s)$$

gilt.

Wir werden später beim Beweis eines Satzes davon Gebrauch machen, daß die "physikalische" Serien- und Parallelschaltung von Automaten wieder zu Automaten führt (und nicht etwa aus dem Bereich der endlichen Automaten herausführt).

Definition 1.5

Ein stochastischer Automat $\mathcal{A} = (X, Y, Z, w, \mu, \pi_0)$ heißt insbesondere Quelle, wenn $|X| = 1$ gilt.

Die Quellen dienen uns dazu, einen "Informationsstrom" herzustellen, damit wir unsere Automaten mit wahrscheinlichkeitstheoretisch beschreibbaren Zeichenketten "speisen" können.

Wenn $\mathcal{A} = (X, Y, Z, w, \mu, \pi)$ insbesondere Quelle ist, so kann die Angabe des einelementigen Eingabealphabetes X entfallen, ohne daß dabei für unsere Zwecke Wesentliches verloren geht. Eine Quelle kann also als ein 5-Tupel der Gestalt

$$Q = (Y, Z, w, \mu, \pi_0)$$

aufgefaßt werden, wobei w jetzt eine bedingte W -Funktion der Gestalt

$$w: (Z) \times (Z) \longrightarrow [0, 1]$$

sein muß. Die Quellenwahrscheinlichkeit w_Q bestimmt sich gemäß

$$w_Q(y^1 y^2 \dots y^s) := \sum_{z^0 \in Z} \pi_0(z^0) \sum_{\substack{z^1 z^2 \dots z^s \in Z^s \\ \wedge \mu(z^i) = y^i \ (i=1 \dots s)}} \prod_{i=0}^{s-1} w(z^{i+1} | z^i) .$$

Falls $\ell(\bar{y}) = s$, so gibt $w_Q(\bar{y})$ die (genau genommen bedingte) Wahrscheinlichkeit dafür an, daß die Quelle - gestartet in der initialen Zustandsverteilung - nach s Takten genau die Ausgabefolge $\bar{y}$ liefert. Es gilt also

$$\sum_{\bar{y} \in Y^s} w_Q(\bar{y}) = 1.$$

Wegen der sequentiellen Arbeitsweise gilt auch für $\bar{y} \in Y^*$

$$\sum_{y \in Y} w_Q(\bar{y}y) = w_Q(\bar{y}).$$

Da bei einer Quelle das Ausgabealphabet Y am wichtigsten ist, wollen wir es Quellenalphabet nennen.

1.3 Matrizenkalkül für Automaten und Quellen

Für die weitere Untersuchung von Automaten bietet sich die Matrizendarstellung als ein sehr geeignetes Hilfsmittel an. Eine strenge Begründung der von uns benutzten Techniken ist nicht erforderlich, da wir uns vollständig auf der Theorie der endlichen Markow-Ketten und den dort gebräuchlichen Techniken abstützen können (vgl. z. B. KEMENY, SNELL [3], DOOB [2]).

Sei $\mathcal{A} = (X, Y, Z, w, \mu, \pi_0)$ ein Automat und sei die Zustandsmenge gemäß

$$Z = \{z_1, z_2, \dots, z_n\}$$

indiziert. Dann ist die Überführungsmatrix A_x , die zum Eingangssignal $x \in X$ gehört, festgelegt gemäß

$$A_x := (a_{xik}) \quad (i=1,2,\dots,n; k=1,2,\dots,n)$$

und

$$a_{xik} := w(z_k | z_i, x)$$

für alle $i, k \in \{1, 2, \dots, n\}$. A_x ist also wegen

$$\sum_{k=1}^n a_{xik} = 1 \quad (i=1,2,\dots,n)$$

eine stochastische n - n -Matrix.

Wir nehmen im folgenden stets an, daß eine Zustandsverteilung π von $\mathcal{U}$ insbesondere als Zeilenvektor

$$\pi = (\pi(z_1), \pi(z_2), \dots, \pi(z_n))$$

aufgefaßt werden kann. Sei $\bar{x} = x^1 x^2 \dots x^s \in X^*$ eine Eingabekette. Dann bezeichnen wir mit $A_{\bar{x}}$ die Matrix

$$A_{\bar{x}} := A_{x^1} \cdot A_{x^2} \cdot \dots \cdot A_{x^s},$$

die wiederum eine stochastische Matrix ist. Sei jetzt π eine beliebige Zustandsverteilung von $\mathcal{U}$ und $\bar{x} \in X^*$. Wenn sich der Automat zu Beginn in der Zustandsverteilung π befindet und wir geben die Eingabekette $\bar{x} = x^1 x^2 \dots x^s$ in den Automaten ein, so befindet er sich anschließend in der Zustandsverteilung π' , wobei

$$\pi' = \pi \cdot A_{x^1} \cdot A_{x^2} \cdot \dots \cdot A_{x^s} = \pi A_{\bar{x}}$$

gilt.

Sei $Y = \{y_1, y_2, \dots, y_m\}$ eine fest gewählte Indizierung der Ausgabesymbole von $\mathcal{U}$. Die $n \times m$ -Matrix

$$B = (b_{ij}) \quad (i=1, 2, \dots, n; j=1, 2, \dots, m)$$

mit

$$b_{ij} := \begin{cases} 1, & \text{falls } \mu(z_i) = y_j, \\ 0 & \text{sonst} \end{cases}$$

heißt Markierungsmatrix von $\mathcal{U}$. Sie hat in jeder Zeile genau eine 1. Wenn π eine Zustandsverteilung von $\mathcal{U}$, so liefert der Vektor

$$\pi B = (v_1, v_2, \dots, v_m)$$

durch die Komponente v_j die Wahrscheinlichkeit dafür, daß der Automat in der Zustandsverteilung π sich in einem Zustand befindet, der mit y_j markiert ist. Wenn B_j den j -ten Spaltenvektor von B bezeichnet, so gilt $v_j = \pi B_j$.

Wenn sich $\mathcal{A}$ zu Beginn in der initialen Verteilung π_0 befindet, und wir geben die Kette $\bar{x} \in X^*$ ein, so ist also

$$\pi_0 A_{\bar{x}} B_j$$

die Wahrscheinlichkeit dafür, daß das letzte von $\mathcal{A}$ ausgegebene Symbol (nach Eingabe von $\bar{x}$) gleich y_j ist.

Die Menge $P(Z)$ bezeichnen wir in naheliegender Weise auch als die Menge der n -dimensionalen -Wahrscheinlichkeitsvektoren. Nur Elemente von $P(Z)$ kommen als Zustandsverteilungen für $\mathcal{A}$ in Betracht. An die Stelle von Z können wir auch n setzen. Wir machen $P(n)$ durch Einführung der Absolutbetrags-Metrik zu einem metrischen Raum: Der Abstand zwischen zwei W-Vektoren

$\pi = (p_1, p_2, \dots, p_n)$ und $\pi' = (p'_1, p'_2, \dots, p'_n)$ sei definiert gemäß

$$|\pi - \pi'| := \sum_{i=1}^n |p_i - p'_i|.$$

Die Multiplikation mit einer stochastischen Matrix ist ein Operator im Raum $P(n)$.

Hilfssatz 1.1

Sei A eine stochastische n - n -Matrix und seien π, π' zwei n -dim. W-Vektoren. Dann gilt

$$|\pi - \pi'| \geq |\pi A - \pi' A|.$$

Beweis: Es gilt

$$\begin{aligned} |\pi A - \pi' A| &= \sum_{j=1}^n \left| \sum_{i=1}^n (p_i a_{ij} - p'_i a_{ij}) \right| \\ &\leq \sum_{j=1}^n \sum_{i=1}^n |p_i - p'_i| \cdot a_{ij} \\ &= \sum_{i=1}^n |p_i - p'_i| \cdot \underbrace{\sum_{j=1}^n a_{ij}}_1 = \sum_{i=1}^n |p_i - p'_i|, \end{aligned}$$

womit der Hilfssatz bewiesen ist.

Stochastische Matrizen sind also schwach kontrahierende Operatoren im Raum $P(n)$. Diese Aussage werden wir später mehrmals ausnutzen. Der Hilfssatz gilt auch für beliebige positive Vektoren π, π' .

2. Vorhersage der Ausgabe eines Automaten

2.1 Die Metrik Δ

Sei Y ein Alphabet. Die Abbildung $\Delta: Y^* \times Y^* \longrightarrow \mathbb{N}$
($\mathbb{N} = \{0, 1, 2, \dots\}$) sei definiert gemäß den Vorschriften:

$$(1) \quad \Delta(y, y') := \begin{cases} 0, & \text{falls } y = y' \\ 1, & \text{falls } y \neq y' \end{cases}$$

für alle Paare $y, y' \in Y$,

$$(2) \quad \Delta(\Lambda, \Lambda) := 0,$$

$$(3) \quad \Delta(y^1 y^2 \dots y^s, y'^1 y'^2 \dots y'^s) := \sum_{i=1}^s \Delta(y^i, y'^i)$$

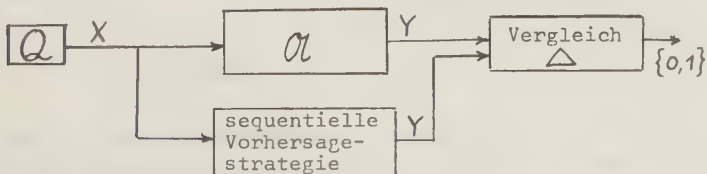
für alle Ketten $\bar{y}, \bar{y}' \in Y^*$ mit $\ell(\bar{y}) = \ell(\bar{y}')$,

$$(4) \quad \Delta(\bar{y}, \bar{y}') := \text{Max} \{ \ell(\bar{y}), \ell(\bar{y}') \}, \text{ falls } \bar{y}, \bar{y}' \in Y^* \text{ und } \ell(\bar{y}) \neq \ell(\bar{y}').$$

Die Abbildung Δ kann als eine verallgemeinerte Hamming-Metrik über der Menge Y^* aufgefaßt werden. Sie wird uns im wesentlichen dazu dienen, die Anzahl der nicht übereinstimmenden Positionen y^i bzw. y'^i in zwei gleichlangen Ketten $\bar{y}, \bar{y}' \in Y^*$ zu messen.

2.2 Die Vorhersageanordnung

Wir haben jetzt alle Vorbereitungen getroffen, um die Vorhersageanordnung vorzustellen und zu diskutieren, die wir in dieser Arbeit untersuchen wollen. Wir betrachten das Diagramm



Gegeben sei also eine Quelle Q sowie ein Automat $\mathcal{A}$. Die Zeichen, die die Quelle liefert, mögen als Eingabe in den Automaten laufen. Der Automat $\mathcal{A}$ reagiert auf die Eingabe mit Ausgabezeichen. Wenn Q z. B. das Ausgabealphabet X und die Quellenwahrscheinlichkeit w_Q hat, so beobachten wir mit Wahrscheinlichkeit $w_Q(\bar{x})$ die Kette $\bar{x} \in X^*$, falls die Quelle genau $\ell(\bar{x})$ Takte läuft, d. h. eine Ausgabekette der Länge $\ell(\bar{x})$ produziert. Die Übertragungsfunktion $w_{\mathcal{A}}$ liefert dann die Aussage, daß mit Wahrscheinlichkeit $w_{\mathcal{A}}(\bar{y}|\bar{x})$ die Ausgabefolge $\bar{y} \in Y^*$ auftritt. Wir verlangen von der sequentiellen Vorhersagestrategie, daß sie in Abhängigkeit von den Eingabesignalen für $\mathcal{A}$ (die von Q geliefert werden) jedes Ausgabesymbol von $\mathcal{A}$ möglichst gut vorhersagt. Um die Güte der Vorhersage zu messen, führen wir die Ausgabesignale von $\mathcal{A}$ sowie die entsprechenden Vorhersagen in einen Vergleichsautomaten Δ , der seinerseits eine 0 ausgibt, falls das tatsächliche Ausgabesignal und das vorhergesagte Ausgabesignal übereinstimmen bzw. eine 1 ausgibt, wenn keine Übereinstimmung vorliegt. Zur formalen Definition des Vorhersagefehlers benötigen wir die

Definition 2.1

Sei X ein Eingabealphabet und Y ein Ausgabealphabet. Eine Vorhersagestrategie von X in Y ist eine bedingte Wahrscheinlichkeitsfunktion

$$V: (Y^*) \times (X^*) \longrightarrow [0,1]$$

mit folgenden Eigenschaften:

$$(1) \quad \sum_{\bar{y} \in Y^*} V(\bar{y}|\bar{x}) = 1 \quad \text{für alle } \bar{x} \in X^*,$$

$$(2) \quad V(\bar{y}|\bar{x}) = 0, \text{ falls } \ell(\bar{y}) \neq \ell(\bar{x}),$$

$$(3) \quad \text{Für alle } \bar{x} \in X^*, \bar{y} \in Y^* \text{ mit } \ell(\bar{x}) = \ell(\bar{y}) \text{ sowie } x \in X \text{ gilt}$$

$$\sum_{y \in Y} V(\bar{y}y|\bar{x}x) = V(\bar{y}|\bar{x})$$

(sequentielle Teilung der Wahrscheinlichkeiten).

Es darf hier - ohne weitergehende Begründung - angemerkt werden, daß jede Vorhersagestrategie V als Übertragungsfunktion eines stochastischen Automaten mit abzählbar unendlich vielen Zuständen aufgefaßt werden kann. Vorhersagestrategien lassen sich im allge-

meinen nicht mit endlichen Automaten realisieren.

Sei Q eine Quelle mit dem Quellenalphabet X und der Quellenwahrscheinlichkeit w_Q . Sei ferner $\mathcal{A} = (X, Y, Z, w, \mu, \pi_0)$ ein Automat mit der Übertragungsfunktion $w_{\mathcal{A}}$ und V eine Vorhersagestrategie von X in Y . Der Erwartungswert für den Vorhersagefehler bei einer Beobachtungsdauer von T Schritten ist dann

$$F(\mathcal{A}, V, Q, T) := \sum_{\bar{x} \in X^T} w_Q(\bar{x}) \sum_{\bar{y}, \bar{y}' \in Y^T} w_{\mathcal{A}}(\bar{y} | \bar{x}) \cdot V(\bar{y}' | \bar{x}) \cdot \Delta(\bar{y}, \bar{y}') .$$

Wir nennen $F(\mathcal{A}, V, Q, T)$ den T -Vorhersagefehler von $\mathcal{A}, V, Q$.

Daß F der Erwartungswert für den Vorhersagefehler ist, kann so eingesehen werden: Unter der Voraussetzung, daß die Quelle in den T Schritten $\bar{x}$ liefert (Quelle und Automat befinden sich zu Beginn in der initialen Zustandsverteilung), so liefert $\mathcal{A}$ mit Wahrscheinlichkeit $w_{\mathcal{A}}(\bar{y} | \bar{x})$ die Kette $\bar{y}$ und - davon unabhängig - V mit Wahrscheinlichkeit $V(\bar{y}' | \bar{x})$ die Vorhersagefolge $\bar{y}'$. Der Vorhersagefehler ist in diesem Fall genau gleich $\Delta(\bar{y}, \bar{y}')$.

Man sieht leicht, daß für den T -Vorhersagefehler stets

$$0 \leq F(\mathcal{A}, V, Q, T) \leq T$$

gilt. Die Beobachtung oder Berechnung des Vorhersagefehlers für eine begrenzte Zeitspanne $1 \dots T$ ist nicht sonderlich interessant. Man kann nämlich V ohne große Mühe so wählen, daß man - vage ausgedrückt - für eine begrenzte Zeitspanne einen recht günstigen Vorhersagefehler erzielt. Es ist jedoch wesentlich schwieriger, auch für beliebig lange Vorhersagezeiten eine im Mittel gute Vorhersage zu erreichen, vorausgesetzt, daß V fest gewählt ist.

Definition 2.2

Der mittlere Vorhersagefehler $\tilde{F}(\mathcal{A}, V, Q)$ für $\mathcal{A}, V$ und Q sei definiert gemäß -

$$\tilde{F}(\mathcal{A}, V, Q) := \lim_{T \rightarrow \infty} \frac{1}{T} F(\mathcal{A}, V, Q, T) .$$

$\tilde{F}(\mathcal{A}, V, Q)$ existiert bzw. existiert nicht je nach dem, ob der Limes existiert bzw. nicht existiert.

In direkter Schreibweise haben wir also

$$\widetilde{F}(\mathcal{A}, V, Q) =$$

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{\bar{x} \in X^T} w_Q(\bar{x}) \sum_{\bar{y}, \bar{y}' \in Y^T} w_{\mathcal{A}}(\bar{y}|\bar{x}) \cdot V(\bar{y}'|\bar{x}) \cdot \Delta(\bar{y}, \bar{y}') .$$

Es ist übrigens ganz einfach, Vorhersagestrategien V zu konstruieren, für die der mittlere Vorhersagefehler nicht existiert. Man kann dabei ähnlich vorgehen wie bei der bekannten Konstruktion von unendlich langen 0-1-Folgen, bei denen die relative Häufigkeit von 1 für linear wachsende Abschnitte nicht konvergiert.

Unser nächstes Ziel soll es sein, die Existenz des mittleren Vorhersagefehlers nachzuweisen für alle Vorhersagestrategien, die durch endliche Automaten realisierbar sind.

Definition 2.3

Sei $\mathcal{A} = (X, Y, Z, w, \mu, \pi_0)$ ein Automat mit der Übertragungsfunktion w . Die mit $\mathcal{A}$ realisierbare Vorhersagestrategie $V_{\mathcal{A}}$ von X in Y ist definiert gemäß

$$V_{\mathcal{A}}(\bar{y}|\bar{x}) := w(\bar{y}|\bar{x}) \text{ für alle } \bar{x} \in X^*, \bar{y} \in Y^*$$

Umgekehrt heißt eine Vorhersagestrategie V mit einem (endlichen) Automaten realisierbar, wenn es einen Automaten $\mathcal{A}$ gibt mit

$$V = V_{\mathcal{A}} .$$

Man weist leicht nach, daß ein so definiertes $V_{\mathcal{A}}$ wirklich eine Vorhersagestrategie nach Def. 2.1 ist.

Hilfssatz 2.1

Sei A eine stochastische n - n -Matrix. Dann existiert eine stochastische n - n -Matrix $\widetilde{A}$, so daß

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{i=1}^T A^i = \widetilde{A} .$$

(A^i = A i -mal mit sich selbst multipliziert)

Den Beweis findet man z. B in DOOB [2].

Satz 2.1

Sei $Q = (X, Z, w, \mu, \pi_0)$ eine Quelle. Dann existiert für fest gewähltes $x_j \in X$

$$\tilde{E}(x_j) = 1 = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{\bar{x} \in X^T} w_Q(\bar{x}) \cdot \Delta(\bar{x}, x_j^T) \quad (x_j^T = \underbrace{x_j x_j \dots x_j}_{T \text{ mal}}),$$

d. h. es existiert der mittlere Erwartungswert für das Auftreten von x_j als Ausgabe der Quelle.

Beweis: Sei A die Überführungsmatrix der Quelle, B die Markierungsmatrix und B_j der j -te Spaltenvektor von B , der zu x_j gehört. Dann gibt

$$\pi_0 A^i B_j$$

die Wahrscheinlichkeit dafür an, daß die Quelle im i -ten Takt genau das Symbol x_j ausgibt. Es gilt also

$$\begin{aligned} \tilde{E}(x_j) &= \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{i=1}^T \pi_0 A^i B_j \\ &= \lim_{T \rightarrow \infty} \pi_0 \left(\frac{1}{T} \sum_{i=1}^T A^i \right) B_j, \end{aligned}$$

und der Limes in der letzten Zeile existiert wegen Hilfssatz 2.1, womit der Satz bewiesen ist.

Satz 2.2

Sei $\mathcal{A}$ ein beliebiger Automat, Q eine beliebige Quelle und V eine Vorhersagestrategie, die mit einem Automaten realisierbar ist. Die Alphabete bzw. Signalmengen seien passend, d. h.

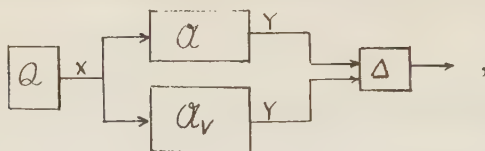
$$\begin{aligned} \mathcal{A} &= (X, Y, Z, w, \mu, \pi_0), \\ \mathcal{Q} &= (X, Z', w_Q, \mu_Q, \pi_Q), \quad \text{und} \end{aligned}$$

V Vorhersagestrategie von X in Y .

Dann existiert

$$\tilde{F}(\mathcal{A}, V, Q).$$

Beweis: Die Vorhersageanordnung für $\tilde{F}(\mathcal{A}, V, Q)$ hat die Gestalt



wobei $\mathcal{A}_V$ ein Automat ist, der V realisiert. $Q, \mathcal{A}, \mathcal{A}_V$ sind Automaten. Der Kasten mit Δ kann ebenfalls als ein Automat dargestellt werden:

$$\Delta = (Y \times Y, \{0, 1\}, \{z_0, z_1\}, w_\Delta, \mu_\Delta, \pi_\Delta).$$

Die restlichen Bestimmungsstücke von Δ werden bestimmt durch

$$\mu_\Delta(z_0) := 0,$$

$$\mu_\Delta(z_1) := 1,$$

für alle $z \in \{z_0, z_1\}$ soll

$$w_\Delta(z_0 | z, (y, y')) := \begin{cases} 1, & \text{falls } y = y' \\ 0 & \text{sonst} \end{cases}$$

sowie

$$w_\Delta(z_1 | z, (y, y')) := \begin{cases} 1, & \text{falls } y \neq y' \\ 0 & \text{sonst} \end{cases}$$

gelten. Der Automat Δ liefert als Ausgabesignale nur 0 oder 1.

π_Δ hat auf die Ausgabe keinen Einfluß, braucht also nicht festgelegt zu werden.

Die gesamte Vorhersageanordnung kann jetzt als eine Quelle Q' mit dem Quellenalphabet $\{0, 1\}$ aufgefaßt werden. Denn ersetze aus formalen Gründen das Quellenalphabet X von Q durch $X \times X$ und wähle μ'_Q gemäß $\mu'_Q(z) := (\mu_Q(z), \mu_Q(z))$ für alle $z \in Z'$. Dann gilt

$$Q' = Q \longrightarrow (\mathcal{A} \parallel \mathcal{A}_V \longrightarrow \Delta)$$

(vgl. Serien- und Parallelschaltung in Abschnitt 1.2) und Q' ist Quelle. Es gilt offensichtlich, daß $F(\mathcal{A}, V, Q, T)$ gleich ist dem Erwartungswert

$$E_Q, (1, T) := \sum_{\bar{y} \in \{0, 1\}^T} w_Q, (\bar{y}) \cdot \Delta(\bar{y}, O^T)$$

für die Quelle Q' . Da aber nach Satz 2.1

$$\tilde{E}(1) = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{i=1}^T E_Q, (1, T)$$

existiert, existiert auch $\tilde{F}(\mathcal{A}, V, Q)$, womit der Satz bewiesen ist.

Damit haben wir ein erstes Ziel unserer Untersuchung erreicht. Für alle durch endliche Automaten realisierbaren Vorhersagestrategien V existiert der mittlere Vorhersagefehler $\tilde{F}(\mathcal{A}, V, Q)$.

Definition 2.4

Sei Q eine Quelle und $\mathcal{A}$ ein Automat.

$$\tilde{F}_{\text{stoch}}(\mathcal{A}, Q) := \inf_{\mathcal{A}' \text{ Automat}} \tilde{F}(\mathcal{A}, V, \mathcal{A}', Q)$$

$$\tilde{F}_{\text{min}}(\mathcal{A}, Q) := \inf_{V \text{ Vorhersagestrategie}} \tilde{F}(\mathcal{A}, V, Q)$$

Dabei ist das Infimum in der letzten Zeile nur über solche V zu erstrecken, für die $\tilde{F}(\mathcal{A}, V, Q)$ existiert. (Wir setzen hier und im folgenden stillschweigend voraus, daß die Alphabete von $\mathcal{A}$, V und Q in ersichtlicher Weise zueinander passen, wenn wir einen Fehler $F(\mathcal{A}, V, Q, T)$, $\tilde{F}(\mathcal{A}, V, Q)$, $\tilde{F}_{\text{stoch}}(\mathcal{A}, Q)$ oder $\tilde{F}_{\text{min}}(\mathcal{A}, Q)$ diskutieren.)

Es gilt offensichtlich stets

$$\tilde{F}_{\text{stoch}}(\mathcal{A}, Q) \geq \tilde{F}_{\text{min}}(\mathcal{A}, Q),$$

denn die durch Automaten realisierbaren Vorhersagestrategien sind insbesondere Vorhersagestrategien. $\tilde{F}_{\text{min}}$ gibt also an, wie klein der Vorhersagefehler gemacht werden kann, wenn beliebige Vorhersagestrategien verwendet werden dürfen, während $\tilde{F}_{\text{stoch}}$ angibt, wie klein der Fehler gemacht werden kann, wenn nur Vorhersagestrategien benutzt werden dürfen, die mit Automaten realisierbar sind.

Es kann durch einfache automaten-theoretische Überlegungen nachgewiesen werden, daß stochastische Automaten mit unendlich vielen Zuständen in bezug auf ihre Übertragungsfunktionen wesentlich leistungsfähiger sind als endliche stochastische Automaten. Dies gilt insbesondere bei Fragestellungen, bei denen das "Langzeitverhalten" eine Rolle spielt. Bereits mit deterministischen erkennenden Automaten, die abzählbar unendlich viele Zustände haben, läßt sich jede beliebige formale Sprache $LC X^*$ erfassen, was z. B. ganz wesentlich über die Leistungsfähigkeit der Turingmaschine hinausreicht.

Trotz dieses deutlichen "Leistungsüberhanges", der sich natürlich auch auf allgemeine Vorhersagestrategien überträgt, werden wir nachweisen können, daß stets

$$\tilde{F}_{\text{stoch}}(\mathcal{A}, Q) = \tilde{F}_{\text{min}}(\mathcal{A}, Q)$$

gilt. Dieser Nachweis ist das Ziel unserer weiteren Untersuchung.

3. Optimale Vorhersage eines Automaten

3.1 Die optimale Vorhersagestrategie V_{opt} für einen Automaten

Eine durch 0 und 1 beschränkte Menge reeller Zahlen hat stets ein Infimum. Daher existiert $\tilde{F}_{\min}(\mathcal{A}, Q)$ ohne weitere ins Detail gehende Überlegungen. Damit ist jedoch noch nicht gesagt, daß es für jede Kombination $\mathcal{A}$, Q auch eine Vorhersagestrategie V gibt mit

$$\tilde{F}_{\min}(\mathcal{A}, Q) = \tilde{F}(\mathcal{A}, V, Q).$$

Wir werden im folgenden beweisen können, daß dies jedoch allgemein der Fall ist. Wir werden zeigen, daß zu jedem $\mathcal{A}$ eine optimale Vorhersagestrategie V_{opt} existiert, so daß - unabhängig von Q -

$$\tilde{F}(\mathcal{A}, V_{\text{opt}}, Q) = \tilde{F}_{\min}(\mathcal{A}, Q)$$

gilt. Die wesentliche Schwierigkeit liegt dabei nicht darin, ein geeignetes V_{opt} zu finden, sondern für V_{opt} die Existenz von $\tilde{F}(\mathcal{A}, V_{\text{opt}}, Q)$ nachzuweisen.

Sei $\mathcal{A} = (X, Y, Z, w, \mu, \pi_0)$ ein Automat, der für den gesamten Abschnitt 3 beliebig, aber fest vorgegeben sei. Wir definieren für $\bar{x} \in X^*$ mit $\bar{x} = x^1 x^2 \dots x^s$

$$v(j|\bar{x}) := \pi_0 A_{\bar{x}} B_j, \quad ,$$

wobei wir die Terminologie von Abschnitt 1.3 benutzen. $v(j|\bar{x})$ gibt die Wahrscheinlichkeit dafür an, daß das letzte Symbol der Ausgabe von $\mathcal{A}$ bei der Eingabe $\bar{x}$ gleich $y_j \in Y$ ist. Statt $v(j|\bar{x})$ schreiben wir gelegentlich auch $v(y_j|\bar{x})$ oder lassen den Index j an y ganz weg.

Weiter sei

$$Y_{\bar{x}} := \{y \mid y \in Y \wedge v(y|\bar{x}) = \max_{y' \in Y} v(y'|\bar{x})\}.$$

Definition 3.1

Die Vorhersagestrategie V_{opt} für $\mathcal{A}$ definieren wir rekursiv durch

$$V_{\text{opt}}(\Lambda / \Lambda) := 1,$$

$$V_{\text{opt}}(y^1 \dots y^{s-1} y^s | x^1 \dots x^{s-1} x^s) :=$$

$$\begin{cases} V_{\text{opt}}(y^1 \dots y^{s-1} | x^1 \dots x^{s-1}) \cdot \frac{1}{|\bar{Y}_{\bar{x}}|}, & \text{falls } y^s \in \bar{Y}_{\bar{x}}, \\ & \text{wobei } \bar{x} = x^1 x^2 \dots x^s, \\ 0 & \text{sonst.} \end{cases}$$

V_{opt} ist also eine Vorhersagestrategie von X in Y .

V_{opt} ist im Sinne unserer Definition 2.1 eine Vorhersagestrategie, denn man verifiziert sofort

$$(1) V_{\text{opt}}(\Lambda / \Lambda) = 1.$$

Sei $\sum_{\bar{y} \in Y^*} V_{\text{opt}}(\bar{y} | \bar{x}) = 1$, so folgt

$$\sum_{\bar{y} y \in Y^*} V_{\text{opt}}(\bar{y} y | \bar{x} x) = \sum_{\bar{y} \in Y^*} V_{\text{opt}}(\bar{y} | \bar{x}) \cdot \sum_{y \in \bar{Y}_{\bar{x}x}} \frac{1}{|\bar{Y}_{\bar{x}x}|} = 1.$$

$$(2) V_{\text{opt}}(\bar{y} | \bar{x}) = 0, \text{ falls } \ell(\bar{y}) \neq \ell(\bar{x}).$$

$$(3) \sum_{y \in Y} V_{\text{opt}}(\bar{y} y | \bar{x} x) = V_{\text{opt}}(\bar{y} | \bar{x}) \cdot \sum_{y \in \bar{Y}_{\bar{x}x}} \frac{1}{|\bar{Y}_{\bar{x}x}|} = V_{\text{opt}}(\bar{y} | \bar{x})$$

(sequentielle Teilung der Wahrscheinlichkeiten).

Es gilt offensichtlich $V_{\text{opt}}(\bar{y} | \bar{x}) = V_{\text{opt}}(\bar{y}' | \bar{x})$, falls $V_{\text{opt}}(\bar{y} | \bar{x}) > 0$ und $V_{\text{opt}}(\bar{y}' | \bar{x}) > 0$. Sei jetzt für $\bar{x} = x^1 x^2 \dots x^s$ und $\bar{y} = y^1 y^2 \dots y^s$

$$E(\Delta(\bar{y}, \cdot) | \bar{x}) := \sum_{\bar{y}' \in Y^s} w_{\bar{x}}(\bar{y}' | \bar{x}) \cdot \Delta(\bar{y}', \bar{y})$$

der Erwartungswert für den Teil-Vorhersagefehler unter der Voraussetzung, daß $\bar{x}$ in den Automaten $\mathcal{A}$ eingegeben wird und daß $\bar{y}$ als Ausgabe vorhergesagt wird.

Satz 3.1

Es gilt für $\mathcal{A}$ und $\bar{x} = x^1 x^2 \dots x^s \in X^s$:

$$V_{\text{opt}}(\bar{y}|\bar{x}) > 0 \iff E(\Delta(\bar{y}, \cdot) | \bar{x}) = \min_{\bar{y}' \in Y^s} E(\Delta(\bar{y}', \cdot) | \bar{x})$$

Beweis: Für $\bar{x} = x^1 x^2 \dots x^s$ und $\bar{y} = y^1 y^2 \dots y^s$ gilt

$$\begin{aligned} E(\Delta(\bar{y}, \cdot) | \bar{x}) &= \sum_{\bar{y}' \in Y^s} w(\bar{y}' | \bar{x}) \cdot \Delta(\bar{y}', \bar{y}) \\ &= \sum_{\bar{y}' \in Y^s} w(\bar{y}' | \bar{x}) \cdot \sum_{i=1}^s \Delta(y^i, y'^i) \\ &= \sum_{i=1}^s \sum_{\bar{y}' \in Y^s} w(\bar{y}' | \bar{x}) \cdot \Delta(y^i, y'^i) \\ &= s - \sum_{i=1}^s v(y^i | x^1 x^2 \dots x^i), \end{aligned}$$

wegen

$$\sum_{\bar{y}_1 \in Y^{i-1} \wedge \bar{y}_2 \in Y^{s-i}} w(\bar{y}_1 y^i \bar{y}_2 | \bar{x}) = \sum_{\bar{y}_1 \in Y^{i-1}} w(\bar{y}_1 y^i | x^1 \dots x^i).$$

$E(\Delta(\bar{y}, \cdot) | \bar{x})$ ist also genau dann minimal, wenn

$$v(y^i | x^1 x^2 \dots x^i) = \max_y v(y | x^1 x^2 \dots x^i)$$

für $i = 1, 2, \dots, s$. Dies ist dann und nur dann der Fall, wenn $V_{\text{opt}}(\bar{y}|\bar{x}) > 0$, was aus der Definition von V_{opt} sofort abgelesen werden kann.

Satz 3.2

Sei $\mathcal{A}$ ein Automat, Q eine Quelle, V_{opt} die optimale Vorhersagestrategie für $\mathcal{A}$. Dann gilt für eine beliebige Vorhersagestrategie V stets

$$F(\mathcal{A}, V, Q, T) \geq F(\mathcal{A}, V_{\text{opt}}, Q, T).$$

Beweis:

$$\begin{aligned} F(\mathcal{A}, V_{\text{opt}}, Q, T) &= \sum_{\bar{x} \in X^T} w_Q(\bar{x}) \sum_{\bar{y}, \bar{y}' \in Y^T} w_{\mathcal{A}}(\bar{y}' | \bar{x}) V_{\text{opt}}(\bar{y} | \bar{x}) \cdot \Delta(\bar{y}, \bar{y}') \\ &= \sum_{\bar{x} \in X^T} w_Q(\bar{x}) \sum_{\bar{y} \in Y^T} V_{\text{opt}}(\bar{y} | \bar{x}) \sum_{\bar{y}' \in Y^T} w_{\mathcal{A}}(\bar{y}' | \bar{x}) \cdot \Delta(\bar{y}, \bar{y}') \\ &= \sum_{\bar{x} \in X^T} w_Q(\bar{x}) \sum_{\bar{y} \in Y^T} V_{\text{opt}}(\bar{y} | \bar{x}) \cdot E(\Delta(\bar{y}, \cdot) | \bar{x}) \\ &\leq \sum_{\bar{x} \in X^T} w_Q(\bar{x}) \sum_{\bar{y} \in Y^T} V(\bar{y} | \bar{x}) \cdot E(\Delta(\bar{y}, \cdot) | \bar{x}) \end{aligned}$$

wegen Satz 3.1 und wegen

$$\sum_{\bar{y} \in Y^T} V_{\text{opt}}(\bar{y} | \bar{x}) = \sum_{\bar{y} \in Y^T} V(\bar{y} | \bar{x}) = 1.$$

Damit haben wir bereits die Aussage

$$\tilde{F}(\mathcal{A}, V_{\text{opt}}, Q) = \tilde{F}_{\min}(\mathcal{A}, Q),$$

falls $\tilde{F}(\mathcal{A}, V_{\text{opt}}, Q)$ existiert!

3.2 Der Raum der Dichten

R_+ bezeichne die positiven reellen Zahlen mit Einschluß der 0.

Seien $\pi = (p_1, p_2, \dots, p_n)$ und $\pi' = (p'_1, p'_2, \dots, p'_n)$ zwei Vektoren $\pi, \pi' \in R_+^n$.

$$\begin{aligned} |\pi| &:= \sum_{i=1}^n |p_i|, \\ |\pi - \pi'| &:= \sum_{i=1}^n |p_i - p'_i|, \end{aligned}$$

$\alpha \pi := (\alpha p_1, \alpha p_2, \dots, \alpha p_n)$, wobei $\alpha > 0$ reell,

$\pi + \pi' := (p_1 + p'_1, p_2 + p'_2, \dots, p_n + p'_n)$.

π heißt parallel zu π' , geschrieben $\pi \parallel \pi'$, wenn ein $\alpha > 0$ existiert mit $\alpha \cdot \pi = \pi'$.

Eine Folge (π_i) ($i=1,2,\dots$) von Vektoren $\pi_i \in \mathbb{R}_+^n$ heißt n -dimensionale Dichte, wenn nur für endlich viele der π_i $|\pi_i| > 0$ gilt.

$$M(\pi_i) := \sum_{i=1}^{\infty} |\pi_i|$$

heißt die Masse von (π_i) .

Zwei Dichten (π_i) und (π'_i) heißen äquivalent, geschrieben

$(\pi_i) \cong (\pi'_i)$, wenn für jeden Vektor $\pi_0 \in \mathbb{R}_+^n$

$$\sum_{\pi_i \parallel \pi_0} |\pi_i| = \sum_{\pi'_i \parallel \pi_0} |\pi'_i|$$

gilt.

Die Operationen (1), (2), (3), auf eine Dichte (π_i) angewandt, führen wieder zu Dichten, die zu (π_i) äquivalent sind:

- (1) Permutierung der Indizes (Umordnungen).
- (2) Ersetzen eines Elementes π_i durch die zwei Elemente $\alpha \cdot \pi_i$ und $(1-\alpha) \cdot \pi_i$ ($0 \leq \alpha \leq 1$), dabei erhält $\alpha \cdot \pi_i$ den Index i , $(1-\alpha) \pi_i$ den Index $i+1$ und π_j mit $j > i$ erhält den Index $j+1$ (Verfeinerung von Dichten).
- (3) Ersetzen zweier Glieder π_i, π_{i+1} durch $\pi_i + \pi_{i+1}$, falls $\pi_i \parallel \pi_{i+1}$. Das neue Glied erhält den Index i , π_j mit $j > i+1$ erhält den Index $j-1$ (Verschmelzung paralleler Glieder).

$$D(\pi_i) := \{ (\pi'_i) \mid (\pi'_i) \cong (\pi_i) \}$$

bezeichnet die Äquivalenzklasse von Dichten, in der (π_i) liegt.

Wir bezeichnen von jetzt ab solche Klassen D als Dichten und

$(\pi_i) \in D$ als einen Repräsentanten von D . Seien (π_i) und (π'_i) beliebige Repräsentanten.

$$|(\pi_i) - (\pi'_i)| := \sum_{i=1}^{\infty} |\pi_i - \pi'_i|.$$

Die Masse $M(D) := M(\pi_i)$, falls $(\pi_i) \in D$. Diese Definition ist sinnvoll, denn äquivalente Repräsentanten haben stets die gleiche Masse. Die Dimension von D ist definiert durch die Dimension der Repräsentanten.

Definition 3.2

Seien D, D' Dichten gleicher Dimension.

$$\mathcal{S}(D, D') := \inf_{\substack{(\pi_i) \in D \\ (\pi'_i) \in D'}} |(\pi_i) - (\pi'_i)|.$$

Satz 3.3

$\mathcal{S}$ ist über der Menge der Dichten gleicher Dimension eine Metrik.

Beweis: (1) Die Symmetrie $\mathcal{S}(D, D') = \mathcal{S}(D', D)$ ergibt sich unmittelbar aus der Definition.

(2) Wir beweisen jetzt $\mathcal{S}(D, D') = 0 \Rightarrow D = D'$. $\bar{\pi}$ heißt Hauptvektor von D , wenn für $(\pi_i) \in D$

$$|\bar{\pi}| = \sum_{\pi_i \parallel \bar{\pi}} |\pi_i| > 0.$$

Die endlich vielen Hauptvektoren von D charakterisieren D vollständig. Seien $\bar{\pi}_j$ ($j=1, \dots, m$) bzw. $\bar{\pi}'_k$ ($k=1, \dots, m'$) die Hauptvektoren von D bzw. D' . Seien $(\pi_i) \in D$ und $(\pi'_i) \in D'$. Jeder Vektor π_i mit $|\pi_i| > 0$ ist parallel zu genau einem der Hauptvektoren von D . Definiere für $j=1, 2, \dots, m$ und $k=1, 2, \dots, m'$

$$\alpha_{jk} := \frac{1}{|\bar{\pi}_j|} \sum_{\substack{\pi_i \parallel \bar{\pi}_j \\ \wedge \pi_i \parallel \bar{\pi}'_k}} |\pi_i|, \quad \beta_{jk} := \frac{1}{|\bar{\pi}'_k|} \sum_{\substack{\pi_i \parallel \bar{\pi}_j \\ \wedge \pi_i \parallel \bar{\pi}'_k}} |\pi_i|.$$

Es gilt jetzt

$$\sum_{\pi_i \parallel \bar{\pi}_j \wedge \pi'_i \parallel \bar{\pi}'_k} |\pi_i - \pi'_i| \geq |\bar{\pi}_j \cdot \alpha_{jk} - \bar{\pi}'_k \cdot \beta_{jk}|,$$

also

$$\sum_{j=1}^m \sum_{k=1}^{m'} |\bar{\pi}_j \cdot \alpha_{jk} - \bar{\pi}'_k \cdot \beta_{jk}| \leq |(\pi_i) - (\pi'_i)|.$$

Seien $(\pi_i^v) \in D$, $(\pi'_i{}^v) \in D'$ für $v = 1, 2, 3, \dots$ Repräsentanten, so daß

$$\lim_{v \rightarrow \infty} |(\pi_i^v) - (\pi'_i{}^v)| = 0.$$

Folgen solcher Repräsentanten gibt es genau dann, wenn $\mathcal{S}(D, D') = 0$. Für die Zahlen α_{jk}^v und β_{jk}^v gilt dann auch

$$\lim_{v \rightarrow \infty} \sum_{j=1}^m \sum_{k=1}^{m'} |\bar{\pi}_j \cdot \alpha_{jk}^v - \bar{\pi}'_k \cdot \beta_{jk}^v| = 0.$$

Da $\alpha_{jk}^v \leq 1$ und $\beta_{jk}^v \leq 1$, existiert eine Indexfolge v_s ($s=1, 2, \dots$), so daß

$$\lim_{s \rightarrow \infty} \alpha_{jk}^{v_s} = \bar{\alpha}_{jk}$$

und

$$\lim_{s \rightarrow \infty} \beta_{jk}^{v_s} = \bar{\beta}_{jk},$$

denn die α_{jk}^v und β_{jk}^v haben - als ein Vektor aufgefaßt - mindestens einen Häufungspunkt. Dann gilt aber

$$\sum_{j=1}^m \sum_{k=1}^{m'} |\bar{\pi}_j \bar{\alpha}_{jk} - \bar{\pi}'_k \bar{\beta}_{jk}| = 0,$$

also auch

$$\sum_{k=1}^{m'} |\bar{\pi}_j \bar{\alpha}_{jk} - \bar{\pi}'_k \bar{\beta}_{jk}| = 0$$

und

$$\sum_{k=1}^{m'} \bar{\alpha}_{jk} = 1.$$

Daraus folgt, daß genau einer der Vektoren $\bar{\pi}'_k$ zu $\bar{\pi}_j$ parallel

ist (mehr als einer können es nicht sein), etwa $\overline{\pi}_{k_0}'$, und daß $\beta_{jk} = 0$ für $k \neq k_0$. Weiter folgt $\overline{\alpha}_{jk} = 0$ für $k \neq k_0$, also

$$\overline{\pi}_j = \overline{\beta}_{jk_0} \cdot \overline{\pi}_{k_0}'$$

mit $\overline{\beta}_{jk_0} \leq 1$.

Zu jedem Vektor $\overline{\pi}_j$ gibt es also einen Vektor $\overline{\pi}_{k_0}'$ mit $\overline{\pi}_j \parallel \overline{\pi}_{k_0}'$ und $|\overline{\pi}_j| \leq |\overline{\pi}_{k_0}'|$. Da unsere Betrachtung bei Vertauschung von D und D' ebenfalls funktioniert, gilt schließlich, daß die Hauptvektoren von D und D' übereinstimmen müssen, also daß $D = D'$. Damit ist $\mathcal{S}(D, D') = 0 \Rightarrow D = D'$ für beliebige Dichten D, D' bewiesen, die Hauptvektoren besitzen. Im Falle $\mathcal{S}(D, D') = 0$ und $M(D) = 0$ folgt sofort $M(D') = 0$, also $D = D'$.

(3) Seien jetzt $(\pi_i) \in D$, $(\pi_i') \in D'$, $(\pi_i'') \in D''$ so gewählt, daß für $\varepsilon > 0$

$$\mathcal{S}(D, D') + \mathcal{S}(D', D'') \geq \sum_{i=1}^{\infty} |\pi_i - \pi_i'| + \sum_{i=1}^{\infty} |\pi_i' - \pi_i''| - \varepsilon.$$

Dies ist immer möglich, da Repräsentanten von Dichten beliebig verfeinert werden dürfen. Dann gilt

$$\begin{aligned} \mathcal{S}(D, D') + \mathcal{S}(D', D'') &\geq \sum_{i=1}^{\infty} |\pi_i - \pi_i''| - \varepsilon \\ &\geq \mathcal{S}(D, D'') - \varepsilon. \end{aligned}$$

Da ε in Abhängigkeit von (π_i) , (π_i') , (π_i'') beliebig klein gewählt werden kann, ist eine Ungleichung der Gestalt

$$\mathcal{S}(D, D') + \mathcal{S}(D', D'') < \mathcal{S}(D, D'')$$

nicht möglich.

Damit ist nachgewiesen, daß $\mathcal{S}$ eine Metrik ist. Es sollen jetzt noch einige leicht beweisbare Eigenschaften der Metrik $\mathcal{S}$ nachgetragen werden, die wir später benutzen werden. Wenn $\pi_1, \pi_2, \dots, \pi_m$ die Hauptvektoren von D und $\pi_1', \pi_2', \dots, \pi_m'$ die Hauptvektoren von D' , so sei $D + D'$ definiert durch den Repräsentanten

$$\pi_1, \pi_2, \dots, \pi_m, \pi_1', \pi_2', \dots, \pi_m', 0, 0, \dots$$

(=Folge von Vektoren) und $\alpha \cdot D$ ($\alpha \geq 0$) durch den Repräsentanten

$$\alpha \pi_1, \alpha \pi_2, \dots, \alpha \pi_m, 0, 0, \dots$$

(0 = Nullvektor). Es gilt dann

$$\mathcal{J}(D_1+D, D_2+D) \leq \mathcal{J}(D_1, D_2),$$

$$\mathcal{J}(\alpha D, \alpha D') = \alpha \cdot \mathcal{J}(D, D') \quad (\alpha \geq 0)$$

und

$$|M(D) - M(D')| \leq \mathcal{J}(D, D') \leq M(D) + M(D').$$

Der metrische Raum der Dichten wird vervollständigt (vgl. z. B. LJUSTERNIK, SOBOLEW [4]), das Resultat nennen wir den vollständigen Raum der Dichten. Dichten mit endlich vielen Hauptvektoren - wie sie oben zunächst eingeführt wurden - nennen wir gelegentlich auch diskrete Dichten, um sie von den auch kontinuierlichen Dichten, die durch die Vervollständigung hinzugekommen sind, zu unterscheiden.

Satz 3.4

Die Kugeln $\{D \mid M(D) \leq r \wedge D \text{ Dichte}\}$ sind kompakt.

Beweis: Nach dem Satz von HAUSDORFF haben wir zu zeigen, daß zu jedem $\varepsilon > 0$ ein endliches Gitter G_ε , bestehend aus Punkten D_i , existiert, so daß für beliebiges D mit $M(D) \leq r$

$$\inf_{D_i \in G_\varepsilon} \mathcal{J}(D, D_i) < \varepsilon$$

gilt.

Sei V_ε eine endliche Menge von Vektoren $\pi \in \mathbb{R}_+^n$ mit $|\pi| = 1$, so daß für beliebiges π_0 mit $|\pi_0| = 1$

$$\inf_{\pi \in V_\varepsilon} |\pi_0 - \pi| < \varepsilon.$$

Zu jedem $\varepsilon > 0$ gibt es endliche Mengen V_ε . Sei jetzt D eine Dichte mit $M(D) \leq r$ und $(\pi_i) \in D$ ein Repräsentant. Die Folge

(π_i') sei definiert gemäß

$$\pi_i' := \pi_k \cdot |\pi_i| ,$$

wobei $\pi_k \in V_\varepsilon$ und π_k speziell so gewählt, daß

$$\left| \pi_k - \frac{\pi_i}{|\pi_i|} \right| = \min_{\pi' \in V_\varepsilon} \left| \pi' - \frac{\pi_i}{|\pi_i|} \right| .$$

Im Falle $|\pi_i| = 0$ sei π_i' ebenfalls der 0-Vektor. Falls π_k dadurch nicht eindeutig bestimmt ist, so wird eine beliebige Wahl unter den minimalen π_k getroffen. (π_i') bestimmt eine Dichte D' . Es gilt ($\tilde{\pi}_i \in V_\varepsilon$)

$$\mathcal{S}(D, D') \leq \sum_{i=1}^{\infty} |\pi_i - \pi_i'| = \sum_{i=1}^{\infty} |\pi_i - \tilde{\pi}_i \cdot |\pi_i|| \leq \sum_{i=1}^{\infty} |\pi_i| \cdot \varepsilon ,$$

$$\text{also } \mathcal{S}(D, D') \leq M(D) \cdot \varepsilon \leq r \cdot \varepsilon .$$

Definiere aus D' eine Dichte D'' dadurch, daß zunächst alle Vektoren gleicher Richtung zusammengefaßt werden und die verbleibenden endlich vielen Vektoren (Anzahl beschränkt durch $|V_\varepsilon|$) auf die Längen

$$\frac{j \cdot r}{s} \quad (j=0, 1, 2, \dots, s)$$

gerundet werden. Dann gilt

$$\mathcal{S}(D', D'') < |V_\varepsilon| \cdot \frac{r}{s} ,$$

denn der Rundungsfehler ist pro Vektor höchstens gleich $\frac{r}{2 \cdot s}$. Wir haben also

$$\mathcal{S}(D, D'') \leq \mathcal{S}(D, D') + \mathcal{S}(D', D'') < r \cdot \varepsilon + |V_\varepsilon| \cdot \frac{r}{s} .$$

$r \cdot \varepsilon$ und $|V_\varepsilon| \cdot \frac{r}{s}$ lassen sich durch die Wahl von ε und danach von s beliebig klein machen. Da es für $\varepsilon > 0$ und natürliche s stets nur endlich viele verschiedene Dichten des Typs D'' gibt, ist der Satz bewiesen.

Wir haben damit auch bewiesen, daß jede Folge D_i ($i=1, 2, \dots$) mit $M(D_i) = r$ wenigstens einen Häufungspunkt hat. Diese Aussage werden wir beim Beweis von Satz 3.6 benutzen.

3.3 Q- $\mathcal{A}$ -Dichten und der Taktoperator

Sei $Q = (X, Z_Q, w_Q, \mu_Q, \pi_Q)$ eine Quelle mit der Markierungsmatrix B_Q und der Überführungsmatrix A_Q . Mit $n_1 = |Z_Q|$ ist A_Q also eine stochastische $n_1 \times n_1$ -Matrix. $A_Q(x)$ sei die bedingte Überführungsmatrix von Q unter der Voraussetzung, daß $x \in X$ als Ausgabe beobachtet wird. Wenn $A_Q^{ij}(x)$ das Element in der i -ten Zeile und der j -ten Spalte von $A_Q(x)$ bezeichnet, so gilt

$$A_Q^{ij}(x) := \begin{cases} A_Q^{ij}, & \text{falls } \mu_Q(z_j) = x, \\ 0 & \text{sonst.} \end{cases}$$

Es gilt also

$$A_Q = \sum_{x \in X} A_Q(x) \quad .$$

Die Matrizen $A_Q(x)$ sind im allgemeinen keine stochastischen Matrizen. Befindet sich die Quelle in der initialen Verteilung π_Q und beobachtet man eine Kette $x^1 x^2 \dots x^S$ am Ausgang, so befindet sich Q danach in der Zustandsverteilung

$$\pi_Q(x^1 \dots x^S) := \frac{\pi_Q \cdot A_Q(x^1) A_Q(x^2) \dots A_Q(x^S)}{|\pi_Q \cdot A_Q(x^1) A_Q(x^2) \dots A_Q(x^S)|} \quad ,$$

falls $w_Q(x^1 x^2 \dots x^S) > 0$. Es gilt

$$w_Q(x^1 x^2 \dots x^S) = |\pi_Q A_Q(x^1) A_Q(x^2) \dots A_Q(x^S)| \quad .$$

Wir definieren $A_Q(x^1 x^2 \dots x^S) := A_Q(x^1) A_Q(x^2) \dots A_Q(x^S)$.

Definition 3.3

Sei Q eine Quelle wie oben und $\mathcal{A} = (X, Y, Z, w, \mu, \pi_0)$ ein Automat. Die Q- $\mathcal{A}$ -Dichte D_t nach t Arbeitstakten ist definiert gemäß

$$D_t := \left\{ (\pi', \pi) \mid \pi' = \pi_Q A_Q(x^1 x^2 \dots x^t) \wedge \pi = |\pi'| \cdot \pi_0 A_{x^1 x^2 \dots x^t} \wedge x^1 x^2 \dots x^t \in X^t \right\} .$$

Dabei soll (π', π) einen Vektor mit $|Z_Q| + |Z|$ Komponenten bezeichnen, das Komma separiert die Anteile, die zu Q und $\mathcal{A}$ gehören ($A_{x_1 x_2 \dots x_t}$ Überführungsmatrix von $\mathcal{A}$).

Wir verwenden hier und im folgenden die Mengenschreibweise für Dichten, allerdings mit der Nebenbedingung, daß mehrfach auftretende gleiche Elemente auch mehrfach mitgeführt werden. Wir behandeln Dichten also wie indizierte Mengen, verzichten aber auf explizite Indizierung. Wir unterscheiden auch nicht mehr streng zwischen Dichten und Repräsentanten.

Satz 3.5

Für $f(\mathcal{A}, v_{\text{opt}}, Q, t) := F(\mathcal{A}, v_{\text{opt}}, Q, t) - F(\mathcal{A}, v_{\text{opt}}, Q, t-1)$

und $t \geq 1$ gilt

$$f(\mathcal{A}, v_{\text{opt}}, Q, t) = \sum_{(\pi', \pi) \in D_t} (|\pi| - \text{MaxKomp}(\pi B)) ,$$

wobei $\text{MaxKomp}(p_1, p_2, \dots, p_m) := \max_{i=1}^m p_i$

(B = Markierungsmatrix von $\mathcal{A}$).

Beweis: Es gilt mit $v(y_j | \bar{x}) = \pi_{0\bar{x}} A_{\bar{x}} B_j$

$$f(\mathcal{A}, v_{\text{opt}}, Q, t) = \sum_{\bar{x} \in X^t} w_Q(\bar{x}) \sum_{\substack{\bar{y} \in Y^t \\ \wedge \bar{y}' \in Y^t}} w(\bar{y} | \bar{x}) v_{\text{opt}}(\bar{y}' | \bar{x}) \cdot \Delta(y, y')$$

$$= \sum_{\bar{x} \in X^t} w_Q(\bar{x}) \sum_{\substack{y \in Y \\ \wedge y' \in Y}} \Delta(y, y') \cdot v(y | \bar{x}) \cdot v_{\bar{x}}(y') ,$$

$$\text{denn } v(y | \bar{x}) = \sum_{\bar{y} \in Y^{t-1}} w(\bar{y} | \bar{x}) \quad \text{und}$$

$$v_{\bar{x}}(y') := \sum_{\bar{y}' \in Y^{t-1}} v_{\text{opt}}(\bar{y}' y' | \bar{x}) = \begin{cases} \frac{1}{|Y_{\bar{x}}|} , & \text{falls } y' \in Y_{\bar{x}} \\ 0 & \text{sonst} \end{cases}$$

(vergl. Def. von V_{opt}) und da $\Delta(y, y')$ nicht von $\bar{y}$ und $\bar{y}'$ abhängt, können beide Teilsummierungen durchgeführt werden. Es gilt nun

$$\begin{aligned} \sum_{\substack{y \in Y \\ y' \in Y}} \Delta(y, y') v(y|\bar{x}) V_{\bar{x}}(y') &= \sum_{\substack{y, y' \in Y \\ \wedge y \neq y'}} v(y|\bar{x}) V_{\bar{x}}(y') \\ &= \sum_{y' \in Y} V_{\bar{x}}(y') \cdot \sum_{\substack{y \in Y \\ \wedge y \neq y'}} v(y|\bar{x}) = \sum_{y' \in Y} V_{\bar{x}}(y') \cdot (1 - \text{Max}_Y v(y|\bar{x})) \\ &= (1 - \text{Max}_Y v(y|\bar{x})), \end{aligned}$$

da $V_{\bar{x}}(y') > 0$ genau dann, wenn $v(y'|\bar{x}) = \text{Max}_Y v(y|\bar{x})$.

Wegen $\text{MaxKomp}(\pi_O A_{\bar{x}} B) = \text{Max}_Y v(y|\bar{x})$ haben wir schließlich

$$\begin{aligned} f(\alpha, V_{\text{opt}}, Q, t) &= \sum_{\bar{x} \in X^t} w_Q(\bar{x}) \cdot (1 - \text{MaxKomp}(\pi_O A_{\bar{x}} B)) \\ &= \sum_{(\pi', \pi) \in D_t} (|\pi| - \text{MaxKomp}(\pi B)) , \end{aligned}$$

denn $|\pi'| = |\pi| = w_Q(\bar{x})$, falls $(\pi', \pi) \in D_t$ und

$$\alpha \cdot \text{MaxKomp}(\pi) = \text{MaxKomp}(\alpha \pi) ,$$

womit der Satz bewiesen ist.

Hilfssatz 3.1

Seien $A_Q(x)$ ($x \in X$) die bedingten Überführungsmatrizen von Q , A_Q die Überführungsmatrix von Q . Weiter seien π, π' positive Vektoren mit $|\pi|, |\pi'| \leq 1$. Dann gilt

$$\begin{aligned} \sum_{\bar{x} \in X^S} ||\pi A_Q(\bar{x})| - |\pi' A_Q(\bar{x})|| &\leq \\ \sum_{\bar{x} \in X^S} |\pi A_Q(\bar{x}) - \pi' A_Q(\bar{x})| &\leq |\pi - \pi'| . \end{aligned}$$

Beweis: Für $s = 1$ (Induktionsverankerung) gilt

$$\sum_{x \in X} |\pi_{A_Q}(x) - \pi'_{A_Q}(x)| = |\pi_{A_Q} - \pi'_{A_Q}| \leq |\pi - \pi'|,$$

da $A_Q(x)$ höchstens in den Spalten echt positiv ist, die zu Zuständen gehören, die mit x markiert sind. Stochastische Matrizen sind stets schwach kontrahierend (vgl. Hilfssatz 1.1). Wenn die Behauptung für Längen s gilt, so auch für $s+1$, denn es gilt

$$\begin{aligned} & \sum_{x\bar{x} \in X^{s+1}} |\pi_{A_Q}(x\bar{x}) - \pi'_{A_Q}(x\bar{x})| \\ &= \sum_{x \in X} \sum_{\bar{x} \in X^s} |\pi_{A_Q}(x) \cdot A_Q(\bar{x}) - \pi'_{A_Q}(x) \cdot A_Q(\bar{x})| \\ &\leq \sum_{x \in X} |\pi_{A_Q}(x) - \pi'_{A_Q}(x)| \leq |\pi - \pi'| \end{aligned}$$

unter Zuhilfenahme der Induktionsvoraussetzung.

Für positive Vektoren gilt stets $(\pi_i = (p_{i1}, p_{i2}, \dots, p_{in}), i=1,2)$

$$||\pi_1| - |\pi_2|| = \left| \left(\sum_{i=1}^n p_{1i} \right) - \left(\sum_{i=1}^n p_{2i} \right) \right| \leq \sum_{i=1}^n |p_{1i} - p_{2i}| = |\pi_1 - \pi_2|,$$

womit der Hilfssatz bewiesen ist.

Definition 3.4

Der Taktoperator A für die Quelle Q und den Automaten $\mathcal{A}(n_1$ bzw. n_2 Zustände) wird definiert für einelementige $(n_1+n_2\text{-dim.})$ Q - $\mathcal{A}$ -Dichten $\{(\pi', \pi)\}$ gemäß

$$\{(\pi', \pi)\}_A := \left\{ \left(\pi'_{A_Q}(x), \pi_{A_Q} \cdot \frac{|\pi'_{A_Q}(x)|}{|\pi'|} \right) \mid x \in X \wedge |\pi'| \neq 0 \right\}$$

und für beliebige diskrete Q - $\mathcal{A}$ -Dichten D gemäß

$$DA := \sum_{(\pi', \pi) \in D} \{(\pi', \pi)\}_A$$

(Addition von Dichten siehe Abschnitt 3.2).

A ist an sich nur für Repräsentanten definiert. Man sieht aber sofort, daß A, angewandt auf äquivalente Repräsentanten, wieder zu äquivalenten Repräsentanten führt. Weiter erkennt man, daß A, ausgehend von der Q- $\mathcal{U}$ -Dichte D_t die Q- $\mathcal{U}$ -Dichte D_{t+1} liefert gemäß

$$D_{t+1} = D_t A.$$

Wenn $D_0 = \{(\pi_Q, \pi_{\mathcal{U}})\}$ die initiale Q- $\mathcal{U}$ -Dichte von Q und $\mathcal{U}$ bezeichnet $(\pi_Q, \pi_{\mathcal{U}})$ initiale Zustandsverteilungen von Q und $\mathcal{U}$, so gilt

$$D_t = D_0 A^t,$$

wobei A^t t-malige Anwendung des Operators A bezeichnet. Der Operator ist offenbar distributiv, d. h. es gilt

$$(D + D')A = DA + D'A,$$

sowie massentreu, d. h.

$$M(DA) = M(D).$$

Ein Taktoperator A für eine Quelle Q mit n_1 Zuständen und einen Automaten $\mathcal{U}$ mit n_2 Zuständen nennen wir auch einen n_1 - n_2 -Taktoperator. Jeder n_1 - n_2 -Taktoperator ist auf eine n_1 - n_2 -dimensionale Q- $\mathcal{U}$ -Dichte D anwendbar. Eine Q- $\mathcal{U}$ -Dichte hat stets die Masse $M(D) = 2$.

Hilfssatz 3.2

Sei A ein Taktoperator für Q und $\mathcal{U}$ und seien D, $\bar{D}$ passende diskrete Dichten. Dann gilt

$$\mathcal{S}(DA^S, \bar{D}A^S) \leq 2 \cdot \mathcal{S}(D, \bar{D}).$$

Beweis: Zunächst beweisen wir den Hilfssatz für einpunktige Dichten, also für

$$D = \{(\pi, \pi')\} \quad \text{und} \quad \bar{D} = \{(\bar{\pi}, \bar{\pi}')\},$$

wobei wir nur $|\pi| = |\pi'| \leq 1$ und $|\bar{\pi}| = |\bar{\pi}'| \leq 1$ voraussetzen. Aus der Definition von A (aus Q und $\mathcal{U}$) ergibt sich

$$\{(\pi, \pi')\} A^S = \{(\pi A_{\mathcal{U}}(\bar{x}), \pi' A_{\mathcal{U}}(\bar{x}) \cdot |\pi A_{\mathcal{U}}(\bar{x})|) \mid \bar{x} \in X^S\}.$$

Zur Vereinfachung der Ausdrücke setzen wir

$$w(\bar{x}) := |\pi A_Q(\bar{x})|, \quad \bar{w}(\bar{x}) := |\bar{\pi} A_Q(\bar{x})|,$$

$$\pi_{\bar{x}} := \pi A_Q(\bar{x}), \quad \bar{\pi}_{\bar{x}} := \bar{\pi} A_Q(\bar{x}),$$

$$\pi'_{\bar{x}} := \pi' A_{\bar{x}}, \quad \bar{\pi}'_{\bar{x}} := \bar{\pi}' A_{\bar{x}}.$$

Wir schätzen nun ab:

$$\begin{aligned} \mathcal{S}(DA^S, \bar{D}A^S) &\leq \sum_{\bar{x} \in X^S} \left| (\pi_{\bar{x}}, \pi'_{\bar{x}} w(\bar{x})) - (\bar{\pi}_{\bar{x}}, \bar{\pi}'_{\bar{x}} \bar{w}(\bar{x})) \right| = \\ &= \sum_{\bar{x} \in X^S} \left| \pi_{\bar{x}} - \bar{\pi}_{\bar{x}} \right| + \sum_{\bar{x} \in X^S} \left| \pi'_{\bar{x}} w(\bar{x}) - \bar{\pi}'_{\bar{x}} \bar{w}(\bar{x}) \right| \\ &\leq |\pi - \bar{\pi}| + \sum_{\bar{x} \in X^S} \left| \pi'_{\bar{x}} w(\bar{x}) - \bar{\pi}'_{\bar{x}} \bar{w}(\bar{x}) \right| + \sum_{\bar{x} \in X^S} \left| \pi'_{\bar{x}} \bar{w}(\bar{x}) - \bar{\pi}'_{\bar{x}} w(\bar{x}) \right| \end{aligned}$$

(wegen Hilfssatz 3.1 und Dreiecksungleichung).

a) Weiter gilt wegen $|\pi w - \pi' w'| \leq |w - w'| \cdot |\pi|$

$$\begin{aligned} \sum_{\bar{x} \in X^S} \left| \pi'_{\bar{x}} w(\bar{x}) - \bar{\pi}'_{\bar{x}} \bar{w}(\bar{x}) \right| &\leq \sum_{\bar{x} \in X^S} |\pi'_{\bar{x}}| \cdot |w(\bar{x}) - \bar{w}(\bar{x})| \\ &\leq |\pi - \bar{\pi}|, \end{aligned}$$

da $|\pi'_{\bar{x}}| \leq 1$ und wegen Hilfssatz 3.1.

$$\begin{aligned} \text{b) } \sum_{\bar{x} \in X^S} \left| \pi'_{\bar{x}} \bar{w}(\bar{x}) - \bar{\pi}'_{\bar{x}} w(\bar{x}) \right| &\leq \sum_{\bar{x} \in X^S} \bar{w}(\bar{x}) \cdot \left| \pi'_{\bar{x}} - \bar{\pi}'_{\bar{x}} \right| \\ &\leq \sum_{\bar{x} \in X^S} \bar{w}(\bar{x}) \cdot |\pi' - \bar{\pi}'| \\ &\leq |\pi' - \bar{\pi}'|, \end{aligned}$$

da $\sum_{\bar{x} \in X^S} \bar{w}(\bar{x}) \leq 1$ (wegen $|\bar{\pi}| \leq 1$) und wegen Hilfssatz 1.1.

Also gilt

$$\begin{aligned} \mathcal{S}(DA^S, \bar{D}A^S) &\leq |\pi - \bar{\pi}| + |\pi - \bar{\pi}| + |\pi' - \bar{\pi}'| \\ &\leq 2 (|\pi - \bar{\pi}| + |\pi' - \bar{\pi}'|) \\ &= 2 \cdot \mathcal{S}(D, \bar{D}). \end{aligned}$$

Die Verallgemeinerung auf beliebige diskrete Dichten erfolgt so:

Wenn D durch den Repräsentanten (π_i, π'_i) und $\bar{D}$ durch $(\bar{\pi}_i, \bar{\pi}'_i)$ ($i=1,2,3,\dots$) gegeben ist, und wenn die Folgen bereits so gewählt sind, daß

$$\mathcal{S}(D, \bar{D}) = \sum_{i=1}^{\infty} |(\pi_i, \pi'_i) - (\bar{\pi}_i, \bar{\pi}'_i)| = \varepsilon$$

gilt, so haben wir wegen der Distributivität von A^S

$$\begin{aligned} \mathcal{S}(DA^S, \bar{D}A^S) &\leq \sum_{i=1}^{\infty} \mathcal{S}(\{(\pi_i, \pi'_i)\}A^S, \{(\bar{\pi}_i, \bar{\pi}'_i)\}A^S) \\ &\leq 2 \cdot \sum_{i=1}^{\infty} |(\pi_i, \pi'_i) - (\bar{\pi}_i, \bar{\pi}'_i)| \\ &= 2 \cdot \mathcal{S}(D, \bar{D}) + 2 \cdot \varepsilon \end{aligned}$$

mit beliebig kleinem $\varepsilon > 0$.

Damit ist der Hilfssatz vollständig bewiesen. Die bewiesene Ungleichung gilt für beliebiges s , wodurch eine recht starke Aussage über Taktoperatoren gewonnen ist.

3.4 Die Existenz des mittleren Vorhersagefehlers bei optimaler Vorhersage

Satz 3.6

Sei A ein n_1 - n_2 -Taktoperator und D_0 eine $(n_1$ - n_2 -dim.) Q - $\mathcal{A}$ -Dichte. Dann existiert eine Dichte $\tilde{D}$ mit

$$\tilde{D} = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{i=1}^T D_0 A^i.$$

Beweis: Da $M(D_0) = 2$, gilt auch für die Folge

$$D(T) := \frac{1}{T} \sum_{i=1}^T D_0 A^i \quad (T=1,2,\dots)$$

$M(D(T)) = 2$. Wegen Satz 3.4 hat also $D(T)$ mindestens einen Häufungspunkt $\tilde{D}$. Für eine geeignet gewählte Indexfolge T_ν ($\nu = 1, 2, \dots$) gilt also

$$\tilde{D} = \varliminf_{\nu \rightarrow \infty} D(T_\nu).$$

Behauptung (1): Jeder Häufungspunkt $\tilde{D}$ von $D(T)$ ist Fixpunkt von A , d. h. es gilt $\tilde{D}A = \tilde{D}$.

Denn für $\nu = 1, 2, 3, \dots$ gilt

$$\mathcal{S}(D(T_\nu)A, D(T_\nu)) = \mathcal{S}\left(\frac{1}{T_\nu} \sum_{i=1}^{T_\nu-1} D_0 A^{i+1}, \frac{1}{T_\nu} \sum_{i=1}^{T_\nu-1} D_0 A^i\right) \leq$$

$$\mathcal{S}\left(\frac{1}{T_\nu} D_0 A^{T_\nu+1}, \frac{1}{T_\nu} D_0 A\right) = \frac{1}{T_\nu} \mathcal{S}(D_0 A^{T_\nu+1}, D_0 A)$$

$$\leq \frac{4}{T_\nu}$$

(vergl. Eigenschaften von $\mathcal{S}$),

also

$$\lim_{\nu \rightarrow \infty} D(T_\nu) = \tilde{D} = \lim_{\nu \rightarrow \infty} D(T_\nu)A = \tilde{D}A,$$

denn A ist ein stetiger Operator (vgl. Hilfssatz 3.2).

Behauptung (2): Die Folgen $D(T)$ und

$$D_k(T) := \frac{1}{T} \sum_{i=1}^T D(k) A^i$$

haben die gleichen Häufungspunkte. Die Behauptung gilt für $k = 1$, denn

$$\mathcal{S} \left(\frac{1}{T} \sum_{i=1}^T D_O A^i, \frac{1}{T} \sum_{i=1}^T (D_O A) A^i \right) \subseteq \frac{1}{T} \mathcal{S} (D_O A, D_O A^{T+1}) \subseteq \frac{4}{T}.$$

Da $\lim_{T \rightarrow \infty} \frac{4}{T} = 0$, muß jeder Häufungspunkt der einen Folge auch

Häufungspunkt der anderen Folge sein.

Wenn wir nun für $k \geq 1$

$$\lim_{T \rightarrow \infty} \mathcal{S} (D(T), D_k(T)) = 0$$

voraussetzen (Induktionsvoraussetzung), so können wir folgern:

$$\mathcal{S} (D(T), D_{k+1}(T)) \subseteq \mathcal{S} (D(T), D_k(T)) + \mathcal{S} (D_k(T), D_{k+1}(T)).$$

Weiter gilt

$$\mathcal{S} (D_k(T), D_{k+1}(T)) = \mathcal{S} \left(\frac{1}{T} \sum_{i=1}^T D(k) A^i, \frac{1}{T} \sum_{i=1}^T D(k+1) A^i \right),$$

dabei läßt sich $D_{k+1}(T)$ aufspalten in die Gestalt

$$\begin{aligned} D_{k+1}(T) &= \frac{1}{T} \sum_{i=1}^T \frac{1}{k+1} \sum_{j=1}^{k+1} D_O A^{i+j} = \\ &= \frac{k}{k+1} \cdot \frac{1}{T} \sum_{i=1}^T \frac{1}{k} \sum_{j=1}^k (D_O A^j) A^i + \frac{1}{T(k+1)} \sum_{i=1}^T D_O A^{k+1+i} \\ &= \frac{k}{k+1} D_k(T) + \frac{1}{T(k+1)} \sum_{i=1}^T D_O A^{k+1+i}, \end{aligned}$$

so daß wir

$$\begin{aligned} \mathcal{S} (D_k(T), D_{k+1}(T)) &\subseteq \overbrace{\mathcal{S} \left(\frac{k}{k+1} D_k(T), \frac{k}{k+1} D_k(T) \right)}^{=0} \\ &+ \mathcal{S} \left(\frac{1}{k+1} D_k(T), \frac{1}{T(k+1)} \sum_{i=1}^T D_O A^{k+1+i} \right) \subseteq \end{aligned}$$

$$\leq \frac{1}{k+1} \mathcal{S}(D_k(T), D(T)A^{k+1})$$

$$\leq \frac{1}{k+1} \cdot \mathcal{S}(D_k(T), D(T)) + \frac{1}{k+1} \mathcal{S}(D(T), D(T)A^{k+1})$$

erhalten. Weiter gilt

$$\begin{aligned} \mathcal{S}(D(T), D(T)A^{k+1}) &= \mathcal{S}\left(\frac{1}{T} \sum_{i=1}^T D_O A^i, \frac{1}{T} \sum_{i=1}^T D_O A^{k+1+i}\right) \\ &\leq \mathcal{S}\left(\frac{1}{T} \sum_{i=1}^{k+1} D_O A^i, \frac{1}{T} \sum_{i=1}^{k+1} D_O A^{T+i}\right) \\ &\leq \frac{4(k+1)}{T} . \end{aligned}$$

Wir erhalten also insgesamt

$$\lim_{T \rightarrow \infty} \mathcal{S}(D(T), D_{k+1}(T)) \leq \lim_{T \rightarrow \infty} \left[\left(1 + \frac{1}{k+1}\right) \cdot \mathcal{S}(D(T), D_k(T)) + \frac{4}{T} \right] = 0$$

nach Induktionsvoraussetzung. Die Behauptung (2) gilt also für $k = 1, 2, 3, \dots$

Behauptung (3): Sei $\tilde{D} = \lim_{v \rightarrow \infty} \frac{1}{T_v} \sum_{i=1}^{T_v} D_O A^i$ ein Häufungspunkt

von $D(T)$ und $\mathcal{S}(D(T_v), \tilde{D}) < \varepsilon_v$. Dann haben die Häufungspunkte von

$$D_{T_v}(T) = \frac{1}{T} \sum_{i=1}^T D(T_v) A^i$$

maximal den Abstand $2 \cdot \varepsilon_v$ von den Häufungspunkten von

$$\tilde{D}(T) = \frac{1}{T} \sum_{i=1}^T \tilde{D} A^i .$$

Zum Beweis der Behauptung (3) schätzen wir ab:

$$\mathcal{S}\left(\frac{1}{T} \sum_{i=1}^T \left(\frac{1}{T_v} \sum_{j=1}^{T_v} D_O A^j\right) A^i, \frac{1}{T} \sum_{i=1}^T \tilde{D} A^i\right) \leq$$

$$\begin{aligned}
 &\leq \frac{1}{T} \sum_{i=1}^T \vartheta \left(\frac{1}{T} \sum_{j=1}^{T_y} D_O A^{i+j}, \tilde{D} A^i \right) \\
 &\leq \frac{1}{T} \sum_{i=1}^T 2 \cdot \vartheta \left(\frac{1}{T_y} \sum_{j=1}^{T_y} D_O A^j, \tilde{D} \right) \quad (\text{wegen Hilfssatz 3.2}) \\
 &\leq \frac{1}{T} \cdot 2 \cdot T \cdot \varepsilon_y = 2 \cdot \varepsilon_y.
 \end{aligned}$$

Damit ist Behauptung (3) bewiesen.

Da ein Häufungspunkt $\tilde{D}$ von $D(T)$ Fixpunkt von A ist (Behauptung (1)), gilt

$$\frac{1}{T} \sum_{i=1}^T \tilde{D} A^i = \frac{1}{T} \sum_{i=1}^T \tilde{D} = \tilde{D}.$$

Da die Folgen $D_k(T)$ ($k=1,2,\dots$) die gleichen Häufungspunkte haben wie $D(T)$ (Behauptung (2)), müssen alle Häufungspunkte von $D(T)$ beliebig nahe bei $\tilde{D}$ liegen (Behauptung (3)). Also hat $D(T)$ nur einen Häufungspunkt und ist also konvergent, womit der Satz bewiesen ist.

Mit dem damit erreichten Konvergenzresultat können wir jetzt endlich den nächsten Satz beweisen:

Satz 3.7

Sei Q eine Quelle, $\mathcal{A}$ ein Automat und V_{opt} die optimale Vorhersagestrategie für $\mathcal{A}$. Dann existiert

$$\tilde{F}(\mathcal{A}, V_{\text{opt}}, Q) = \lim_{T \rightarrow \infty} \frac{1}{T} \cdot F(\mathcal{A}, V_{\text{opt}}, Q, T),$$

und für jede andere Vorhersagestrategie V , für die $\tilde{F}(\mathcal{A}, V, Q)$ existiert, gilt

$$\tilde{F}(\mathcal{A}, V, Q) \geq \tilde{F}(\mathcal{A}, V_{\text{opt}}, Q) = \tilde{F}_{\min}(\mathcal{A}, Q).$$

V_{opt} hängt nicht von der speisenden Quelle ab.

Beweis: Wenn A der Taktoperator für Q und $\mathcal{A}$ ist, π die initiale Zustandsverteilung von Q, π' die initiale Verteilung von $\mathcal{A}$, dann konvergiert die Folge

$$D(T) = \frac{1}{T} \sum_{i=1}^T \{(\pi, \pi')\} A^i$$

für $T \rightarrow \infty$ gegen eine Grenzdichte $\tilde{D}$. (Satz 3.6). Nach Satz 3.5 gilt

$$\begin{aligned} F_T &= \frac{F(\mathcal{A}, v_{\text{opt}, Q, T})}{T} = \frac{1}{T} \sum_{t=1}^T \sum_{(\pi', \pi) \in D_t} (|\pi| - \text{MaxKomp}(\pi B)) \\ &= \sum_{(\pi', \pi) \in D(T)} (|\pi| - \text{MaxKomp}(\pi B)) . \end{aligned}$$

Konvergiert mit $D(T)$ auch F_T ?

Es gilt für positive Vektoren stets

$$|\pi_1 - \pi_2| \geq |\text{MaxKomp}(\pi_1 B) - \text{MaxKomp}(\pi_2 B)| .$$

Also gilt für geeignet gewählte Repräsentanten $(\pi_i) \in D$, $(\pi'_i) \in D'$ stets

$$\begin{aligned} \vartheta(D, D') &= \sum_{i=1}^{\infty} |\pi_i - \pi'_i| - \varepsilon \\ &\geq \sum_{i=1}^{\infty} |\text{MaxKomp}(\pi_i B) - \text{MaxKomp}(\pi'_i B)| - \varepsilon \\ &\geq \left| \sum_{i=1}^{\infty} \text{MaxKomp}(\pi_i B) - \sum_{i=1}^{\infty} \text{MaxKomp}(\pi'_i B) \right| - \varepsilon \end{aligned}$$

für beliebig kleines $\varepsilon > 0$. Mit der Konvergenz von $D(T)$ haben wir daher auch die Konvergenz von F_T . Also gilt

$$\lim_{T \rightarrow \infty} \frac{F(\mathcal{A}, v_{\text{opt}, Q, T})}{T} = \tilde{F}(\mathcal{A}, v_{\text{opt}, Q}) ,$$

d. h. der mittlere Vorhersagefehler $\tilde{F}(\mathcal{A}, v_{\text{opt}, Q})$ existiert. Aus Satz 3.2 folgt sofort für beliebiges V

$$\tilde{F}(\mathcal{A}, v_{\text{opt}, Q}) \leq \tilde{F}(\mathcal{A}, v, Q) ,$$

falls $\tilde{F}(\mathcal{A}, V, Q)$ existiert. Also gilt auch

$$\tilde{F}(\mathcal{A}, V_{\text{opt}}, Q) = \tilde{F}_{\min}(\mathcal{A}, Q).$$

Ein Blick auf die Definition von V_{opt} zeigt schließlich, daß die optimale Vorhersagestrategie für Automaten stets unabhängig von der speisenden Quelle definiert ist.

Die oben eingeführte Vorhersagestrategie V_{opt} , die übrigens im allgemeinen nichtdeterministisch ist, weil es Ketten $\bar{x}$ und $\bar{y}$ mit

$$0 < V_{\text{opt}}(\bar{y}|\bar{x}) < 1$$

geben kann, ist nicht die einzig mögliche Strategie, die das Minimum des mittleren Vorhersagefehlers erreicht. Es ist immer möglich, von V_{opt} ausgehend eine Strategie V mit minimalem Vorhersagefehler zu definieren, die deterministisch ist, d. h. für die

$$V(\bar{y}|\bar{x}) \in \{0, 1\}$$

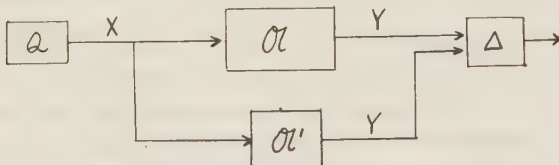
gilt. Jedoch läßt sich eine solche optimale deterministische Strategie nur in den seltensten Fällen von einem endlichen deterministischen Automaten realisieren.

Außerdem hat das "Kurzzeitverhalten" einer Vorhersagestrategie, also die Vorhersage für die Zeiten $1, 2, \dots, T_0 < \infty$ keinen Einfluß auf den sich im Grenzwert ergebenden mittleren Vorhersagefehler.

4. Vorhersage von Automaten mit Automaten

4.1 Multiplex-Technik bei Automaten

In diesem Abschnitt betrachten wir Vorhersageanordnungen der Struktur



für die wir ja bereits bewiesen haben, daß $\tilde{F}(A, v_{A'}, Q)$ stets existiert (vgl. Satz 2.2). Wir setzen uns das Ziel, A' so zu wählen, daß wir einen möglichst geringen Wert für $\tilde{F}(A, v_{A'}, Q)$ erhalten. Wir werden schließlich beweisen können, daß

$$\tilde{F}_{\min}(A, Q) = \tilde{F}_{\text{stoch}}(A, Q),$$

d. h. daß man bereits mit stochastischen Automaten einen gegebenen Automaten A fast beliebig gut vorhersagen kann.

Bevor wir uns näheren Einzelheiten dieses Beweises zuwenden, müssen wir eine Multiplex-Technik für Automaten einführen.

Definition 4.1

Sei $A = (X, Y, Z, w, \mu, \pi_0)$ ein Automat. Der Multiplex-Automat A^M von A wird folgendermaßen definiert:

$$A^M := (X, Y, Z \times Z \times Z, w', \mu', \pi_0'),$$

wobei:

$$a) \quad w'((z_1', z_2', z_3') | (z_1, z_2, z_3), x) := w(z_1' | z_1, x) w(z_2' | z_2, x) w(z_3' | z_3, x)$$

für alle $z_i, z_i' \in Z$ ($i=1,2,3$) und alle $x \in X$.

b) μ' ist eine nichtdeterministische Markierung $\mu': Z^3 \rightarrow P(Y)$, wobei $P(Y)$ die Menge der Wahrscheinlichkeitsverteilungen über

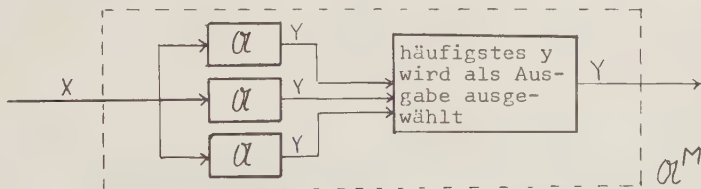
Y bezeichnet. Sei $\mu'_i(z_1, z_2, z_3)$ = Wahrscheinlichkeit dafür, daß $\mathcal{A}^M$ im Zustand $(z_1, z_2, z_3) \in Z^3$ das Signal $y_i \in Y$ ausgibt.

μ'_i sei definiert gemäß

$$\mu'_i(z_1, z_2, z_3) := \begin{cases} 1, & \text{falls wenigstens zwei der Zustände} \\ & z_1, z_2, z_3 \text{ in } \mathcal{A} \text{ mit } y_i \text{ markiert sind,} \\ \frac{1}{3}, & \text{falls alle Zustände } z_1, z_2, z_3 \text{ in } \mathcal{A} \\ & \text{verschieden markiert sind und genau} \\ & \text{einer mit } y_i \text{ markiert ist,} \\ 0 & \text{sonst.} \end{cases}$$

c) $\pi'_0(z_1, z_2, z_3) := \pi_0(z_1) \cdot \pi_0(z_2) \cdot \pi_0(z_3)$ für alle $(z_1, z_2, z_3) \in Z^3$.

Der Automat $\mathcal{A}^M$ hat also folgende Struktur:



$\mathcal{A}^M$ zerfällt also in drei unabhängig voneinander laufende Exemplare von $\mathcal{A}$, wobei die Eingabe in alle drei Exemplare stets identisch ist. In jedem Takt liefert jedes Exemplar ein Ausgabesignal, diese Signale sind im allgemeinen voneinander verschieden. Wenn ein bestimmtes Signal doppelt oder dreifach vorkommt, so wird es als Ausgabesignal für $\mathcal{A}^M$ genommen. Sind alle drei Signale voneinander verschieden, so wird jedes der Signale mit Wahrscheinlichkeit $1/3$ als Ausgabe für $\mathcal{A}^M$ gewählt (kein Mehrheitsentscheid möglich). Wie bereits in Abschnitt 1.1 bemerkt, ist die nichtdeterministische Markierung der Zustände von $\mathcal{A}^M$ keine wesentliche Verallgemeinerung unseres Automatenmodells.

Jeder der drei Automaten liefert das letzte Ausgabesignal nach der Eingabe von $\bar{x} \in X^*$ mit der Verteilung $\pi = \pi_0 A_{\bar{x}}^B$. (Die drei Auto-

maten arbeiten unabhängig voneinander.) Mit welcher Wahrscheinlichkeitsverteilung liefert dann $\mathcal{A}^M$ die Ausgabe?

Hilfssatz 4.1

Seien $\mathcal{A}$, $\mathcal{A}^M$ Automaten wie oben, $Y = \{y_1, y_2, \dots, y_m\}$ das Ausgabealphabet für $\mathcal{A}$ und $\mathcal{A}^M$. Dann gilt:

Wenn $\mathcal{A}$ nach Eingabe von $\bar{x} \in X^*$ die Ausgabeverteilung

$$\pi = (p_1, p_2, \dots, p_m) \quad (= \pi_{\mathcal{O} \wedge \bar{x} B})$$

aufweist, so weist $\mathcal{A}^M$ nach Eingabe von $\bar{x}$ die Ausgabeverteilung

$$\bar{\pi} = (\bar{p}_1, \bar{p}_2, \dots, \bar{p}_m)$$

auf, wobei für $j = 1, 2, \dots, m$

$$\bar{p}_j = p_j \left(1 + p_j - \sum_{i=1}^m p_i^2 \right)$$

gilt.

Beweis: Durch Berücksichtigung der Fallunterscheidungen bei der Auswahl des Ausgabesymbols von $\mathcal{A}^M$ erhält man

$$\bar{p}_j = \underbrace{1 \cdot p_j^3}_{y_j \text{ 3-mal}} + \underbrace{3 \cdot p_j^2 \cdot \sum_{\substack{i=1 \\ i \neq j}}^m p_i}_{y_j \text{ 2-mal}} + \underbrace{\frac{1}{3} \cdot 3 \cdot p_j \cdot \sum_{\substack{i=1 \\ i \neq j}}^m \sum_{\substack{k=1 \\ k \neq i \\ k \neq j}}^m p_i p_k}_{y_j \text{ 1-mal, dabei aber nur mit Wahrscheinlichkeit } 1/3 \text{ ausgewählt}}.$$

Also ergibt sich

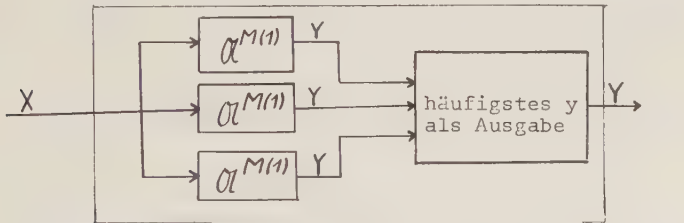
$$\bar{p}_j = p_j^3 + 3 \cdot p_j^2 (1 - p_j) + p_j \left(\sum_{\substack{i=1 \\ i \neq j}}^m p_i (1 - p_i - p_j) \right),$$

wobei wir die Beziehung $1 = \sum_{i=1}^m p_i$ benutzt haben. Wir erhalten

also weiter

$$\begin{aligned} \bar{p}_j &= p_j^3 + 3 \cdot p_j^2(1 - p_j) + p_j \left((1 - p_j) - \sum_{\substack{i=1 \\ i \neq j}}^m p_i^2 - p_j(1 - p_j) \right) \\ &= p_j \left(1 + p_j - \sum_{i=1}^m p_i^2 \right). \end{aligned}$$

Das Prinzip, das bei der Konstruktion des Automaten α^M angewandt wurde, verallgemeinern wir in folgender Weise: Wir bezeichnen mit $\alpha^{M(1)}$ den Automaten α^M , mit $\alpha^{M(2)}$ den Automaten



also $\alpha^{M(2)} = (\alpha^{M(1)})^M$, allgemein also

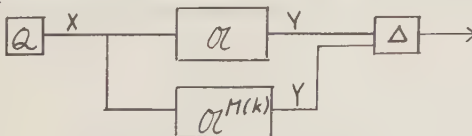
$$\alpha^{M(k+1)} := (\alpha^{M(k)})^M.$$

Wir bezeichnen $\alpha^{M(k)}$ als den k -Multiplex-Automaten von α . In $\alpha^{M(k)}$ kommen also insgesamt 3^k unabhängig voneinander arbeitende Exemplare des Automaten α vor. Allerdings ist $\alpha^{M(k)}$ für natürliche k stets ein (endlicher) Automat.

Ziel der weiteren Untersuchung ist es,

$$\lim_{k \rightarrow \infty} \tilde{F}(\alpha, v_{\alpha^{M(k)}, Q}) = \tilde{F}_{\min}(\alpha, Q) = \tilde{F}_{\text{stoch}}(\alpha, Q)$$

nachzuweisen. Dabei liegt dem mittleren Vorhersagefehler $\tilde{F}(\alpha, v_{\alpha^{M(k)}, Q})$ also eine Vorhersageanordnung der Gestalt



zugrunde.

4.2 Die Ausgabeverteilungen der Multiplex-Automaten

Wir nehmen an, daß α nach der Eingabe von $\bar{x} \in X^*$ in der Ausgabe-verteilung $\pi_{\alpha \bar{x}} = \pi$ weilt. Nach Hilfssatz 4.1 weilt dann $\alpha^{M(1)}$ in der Ausgabeverteilung

$$\pi_j = (p_{11}, p_{12}, \dots, p_{1m}), \text{ wobei}$$

$$p_{1j} = p_j (1 + p_j - \sum_{i=1}^m p_i^2)$$

gilt. Wenn wir mit

$$\pi_k = (p_{k1}, p_{k2}, \dots, p_{km})$$

die Ausgabeverteilung bezeichnen, in der dann $\alpha^{M(k)}$ weilt, so gilt wegen der Konstruktion der $\alpha^{M(k)}$ die Rekursionsformel

$$p_{k+1,j} = p_{kj} (1 + p_{kj} - \sum_{i=1}^m p_{ki}^2) .$$

Den Vektor π_k bezeichnen wir in naheliegender Weise als die k-Iteration von π .

Hilfssatz 4.2

Wenn π eine Ausgabeverteilung, π_k die k-Iteration von π , so existiert eine Grenzverteilung

$$\tilde{\pi} = \lim_{k \rightarrow \infty} \pi_k .$$

Wenn π genau an den Komponenten $i_1, i_2, \dots, i_r$ (Indices) maximal ist, so hat $\tilde{\pi}$ an den Komponenten $i_1, i_2, \dots, i_r$

den Wert $\frac{1}{r}$, die übrigen Komponenten von $\tilde{\pi}$ sind gleich 0.

Beweis: Wir können ohne Beschränkung der Allgemeinheit annehmen, daß die r Komponenten p_i ($i=1,2,\dots,r$) von π maximal sind, d. h.

$$p := p_1 = p_2 = \dots = p_r = \max_i p_i .$$

Dann gibt es ein $\eta > 0$ mit $\eta = \min_{i=r+1}^m (p - p_i) > 0$.

Es gilt (nach einem Iterationsschritt) für π_1 :

$$\begin{aligned} p' &:= p_{11} = p(1 + p - \sum_{j=1}^m p_j^2) \\ &= p + p(p - \sum_{j=1}^m p_j^2) = p + \delta . \end{aligned}$$

Wir wollen den Zuwachs der maximalen Komponenten, also δ , abschätzen. Es gilt

$$\sum_{j=1}^m p_j^2 = \sum_{j=1}^r p_j^2 + \sum_{j=r+1}^m p_j^2 \leq r \cdot p^2 + \sum_{j=r+1}^m (p - \eta) p_j = \dots$$

$$\dots = rp^2 + (p - \eta)(1 - rp) ,$$

$$\text{da } \sum_{j=r+1}^m p_j = 1 - rp .$$

Es gilt also für δ :

$$\begin{aligned} \delta &\geq p(p - rp^2 - (p - \eta)(1 - rp)) = \\ &= p(p - rp^2 - p + \eta + rp^2 - rp\eta) = \\ &= p\eta(1 - rp), \text{ also} \end{aligned}$$

$$\delta \geq p\eta(1 - rp).$$

Da $1 - rp \geq 0$, ist der Zuwachs δ der maximalen Komponenten stets positiv. Er wird nur zu 0, wenn $\eta = 0$, d. h. wenn alle Komponenten von π gleich sind ($r = m$) oder wenn $1 = rp$, d. h. wenn die nicht maximalen Komponenten alle verschwinden. Man sieht außerdem, daß die maximalen Komponenten stärker steigen als die übrigen und daß gleiche Komponenten auch nach dem Iterationsschritt gleich sind. Sei z. B. p_i eine Komponente mit $p \geq p_i$. Dann gilt mit

$$S := \sum_{j=1}^m p_j^2 :$$

$$\bar{p} - \bar{p}_i = p(1 + p - S) - p_i(1 + p_i - S) \geq \dots$$

$$\begin{aligned}
 & \cong (p - p_i) + p(p - S) - p_i(p - S) \\
 & = (p - p_i) + \underbrace{(p - p_i) \cdot (p - S)}_{\geq 0} \\
 & \geq (p - p_i),
 \end{aligned}$$

d. h. der Abstand zwischen den maximalen Komponenten und den nicht maximalen Komponenten kann sich bei der Iteration nur vergrößern. Also konvergiert π_k für $k \rightarrow \infty$ und die restlichen Überlegungen zum vollständigen Beweis des Hilfssatzes sind trivial.

Mit $\tilde{\pi}$ bezeichnen wir den eindeutig bestimmten Vektor

$$\tilde{\pi} = \lim_{k \rightarrow \infty} \pi_k,$$

wobei also π_k die k -Iteration von π ist.

Hilfssatz 4.3

Es gilt, falls $|\tilde{\pi} - \pi| > 0$,

$$\frac{|\tilde{\pi} - \pi_1|}{|\tilde{\pi} - \pi|} \leq 1 - \frac{1}{m} \cdot \delta,$$

wobei

$$\delta = \text{MaxKomp}(\pi) - \sum_{j=1}^m p_j^2.$$

Beweis: Ohne Beschränkung der Allgemeinheit können wir annehmen, daß

$$p := \text{MaxKomp}(\pi) = p_1 = p_2 = \dots = p_r$$

und daß π genau an den r ersten Komponenten maximal ist. Dann gilt, da $\tilde{\pi}$ die Gestalt

$$\tilde{\pi} = \left(\frac{1}{r}, \frac{1}{r}, \dots, \frac{1}{r}, 0, 0, \dots, 0\right)$$

hat,

$$|\tilde{\pi} - \pi| = r\left(\frac{1}{r} - p\right) + (1 - rp) = 2(1 - rp).$$

Wegen

$$p' := \text{MaxKomp}(\pi_1) = p(1 + \delta)$$

erhalten wir

$$|\tilde{\pi} - \pi_1| = 2(1 - rp(1 + \delta)),$$

also

$$\begin{aligned} \frac{|\tilde{\pi} - \pi_1|}{|\tilde{\pi} - \pi|} &= \frac{1 - rp(1 + \delta)}{1 - rp} = 1 - \frac{rp\delta}{1 - rp} \\ &\leq 1 - rp\delta \quad (\text{wegen } 1 - rp \leq 1) \\ &\leq 1 - \frac{1}{m} \cdot \delta \quad (\text{wegen } r \geq 1, p \geq \frac{1}{m}). \end{aligned}$$

Das Produkt von positiven (Zeilen)Vektoren $\pi_i = (p_{i1}, p_{i2}, \dots, p_{im})$ ($i = 1, 2$) definieren wir gemäß

$$\pi_1 \cdot \pi_2 := \sum_{j=1}^m p_{1j} \cdot p_{2j}$$

Hilfssatz 4.4

Wenn π eine Ausgabeverteilung, π_k die k -Iteration von π , so gilt

$$\lim_{k \rightarrow \infty} \pi \cdot \pi_k = \text{MaxKomp}(\pi) = \pi \cdot \tilde{\pi}$$

und $\pi \cdot \pi_{k+1} \geq \pi \cdot \pi_k$ für $k = 1, 2, \dots$, d. h. $\pi \cdot \pi_k$ steigt monoton.

Beweis: Es möge für die Komponenten p_i von π gelten

$$p_{i_1} \geq p_{i_2} \geq \dots \geq p_{i_m}.$$

Wie aus dem Iterationsschema sofort ersichtlich, muß dann auch für die Komponenten von π_k

$$p_{ki_1} \geq p_{ki_2} \geq \dots \geq p_{ki_m}$$

gelten. Es gilt

$$d_{i,j} := p_{k+1,i,j} - p_{k,i,j} \geq 0$$

genau dann, wenn

$$p_{k,i_j} \geq \sum_{i=1}^m p_{k,i}^2$$

(vgl. Iterationsformel). Dann gilt auch

$$\pi \cdot \pi_{k+1} - \pi \cdot \pi_k = \sum_{j=1}^m p_{i_j} \cdot d_{i_j} \geq 0,$$

denn wir haben

$$\sum_{j=1}^m d_{i_j} = 0,$$

und die positiven d_{i_j} haben höhere Gewichte als die negativen.

Damit ist die Monotonie von $\pi \cdot \pi_k$ gesichert, und der Rest ergibt sich durch leichte Rechnung aus Hilfssatz 4.2.

Satz 4.1

Für jedes $\varepsilon > 0$ gibt es eine nur von ε abhängende natürliche Zahl k_ε , so daß für alle (m -dimensionalen) Ausgabeverteilungen π die Beziehung

$$|\pi \cdot \tilde{\pi} - \pi \cdot \pi_{k_\varepsilon}| < \varepsilon$$

gilt. Die Konvergenz von $\pi \pi_k$ gegen $\pi \tilde{\pi} = \text{MaxKomp}(\pi)$ ist also gleichmäßig für alle π .

Beweis: Die Menge

$$P_m := \{ \tilde{\pi} \mid \pi \text{ m-dim. Ausgabeverteilung} \}$$

enthält nur endlich viele Elemente. Aus $\pi \in P_m$ folgt, daß alle von 0 verschiedenen Komponenten von π gleich sind.

Behauptung (1): Sei π_k die k -Iteration von π und

$$|P_m - \pi_k| := \min_{\tilde{\pi} \in P_m} |\tilde{\pi} - \pi_k| < \varepsilon.$$

Dann gilt $|\pi \cdot \tilde{\pi} - \pi \cdot \pi_k| < 2\varepsilon$.

Zum Beweis können wir annehmen, daß $\tilde{\pi} \in P_m$ existiert mit $|\tilde{\pi} - \pi_k| < \varepsilon$ und daß $\tilde{\pi}$ die Gestalt

$$\tilde{\pi} = (\underbrace{\frac{1}{r}, \frac{1}{r}, \dots, \frac{1}{r}}_{r\text{-mal}}, 0, 0, \dots, 0) \quad (r \leq m)$$

hat. Durch Umordnung der Komponenten von π kann diese Situation herbeigeführt werden. Für $\pi_k = (p_{k1}, p_{k2}, \dots, p_{km})$ setzen wir also

$$|\tilde{\pi} - \pi_k| = \sum_{i=1}^r \left| \frac{1}{r} - p_{ki} \right| + \sum_{i=r+1}^m p_{ki} < \varepsilon$$

voraus. Es folgt

$$\left| \frac{1}{r} - p_{ki} \right| < \varepsilon \quad (i=1, 2, \dots, r)$$

und

$$|p_{ki} - p_{kj}| \leq \left| \frac{1}{r} - p_{ki} \right| + \left| \frac{1}{r} - p_{kj} \right| < \varepsilon, \text{ falls } i, j \leq r.$$

Sei $p_k := \text{MaxKomp}(\pi_k)$. Dann gilt

$$|p_k - p_{ki}| < \varepsilon \quad \text{für } i \leq r.$$

Sei $p := \text{MaxKomp}(\pi)$. Dann gilt

$$|p - p_i| < \varepsilon \quad \text{für } i \leq r,$$

denn $|p_k - p_{ki}|$ ist monoton steigend für $k = 1, 2, 3, \dots$ (siehe Abschätzung im Beweis zu Hilfssatz 4.2) und damit folgt aus $|p_k - p_{ki}| < \varepsilon$ auch $|p - p_i| < \varepsilon$.

Damit können wir schließlich abschätzen:

$$\begin{aligned} |\pi \cdot \tilde{\pi} - \pi \cdot \pi_k| &= \text{MaxKomp}(\pi) - \sum_{i=1}^m p_i p_{ki} = p - \sum_{i=1}^m p_i p_{ki} \leq \\ &< p - \sum_{i=1}^r (p - \varepsilon) \left(\frac{1}{r} - \varepsilon \right) - \sum_{i=r+1}^m p_i p_{ki} \\ &\leq p - r(p - \varepsilon) \left(\frac{1}{r} - \varepsilon \right) = \varepsilon(r p + 1 - r \varepsilon) \end{aligned}$$

Wegen $p \leq \frac{1}{r} + \varepsilon$ ergibt sich auch $r p \leq 1 + r \varepsilon$, also erhalten wir

$$|\pi \tilde{\pi} - \pi \pi_k| < \varepsilon(1 + r \varepsilon + 1 - r \varepsilon) = 2 \cdot \varepsilon,$$

womit Behauptung (1) bewiesen ist.

Behauptung (2): Für jedes $\varepsilon > 0$ existiert ein k_ε , so daß für beliebiges π wenigstens ein k_0 mit $0 \leq k_0 \leq k_\varepsilon$ existiert, so daß

$$|p_m - \pi_{k_0}| < \varepsilon$$

gilt. ($\pi_0 = \pi = 0$ -Iteration von π).

Zum Beweis betrachten wir die Funktion

$$\delta(\pi) := \text{MaxKomp}(\pi) - \sum_{i=1}^m p_i^2,$$

die über dem Raum der m-dimensionalen Ausgabevertellungen definiert und dort gleichmäßig stetig ist wegen

$$\begin{aligned} |\delta(\pi_1) - \delta(\pi_2)| &\leq |\text{MaxKomp}(\pi_1) - \text{MaxKomp}(\pi_2)| \\ &\quad + \left| \sum_{i=1}^m p_{2i}^2 - p_{1i}^2 \right| \leq \\ &\leq |\pi_1 - \pi_2| + |\pi_2 - \pi_1| = 2 \cdot |\pi_1 - \pi_2|. \end{aligned}$$

Es gilt stets $\delta(\pi) \geq 0$, und es ist leicht nachzurechnen, daß

$$\delta(\pi) = 0 \iff \pi \in P_m$$

gilt. Auf der abgeschlossenen Menge ($\varepsilon > 0$)

$$P_m(\varepsilon) := \{ \pi \mid |p_m - \pi| \geq \varepsilon \}$$

gilt also

$$\delta_\varepsilon = \min \{ \delta(\pi) \mid \pi \in P_m(\varepsilon) \} > 0.$$

Wegen Hilfssatz 4.3 gilt (auch) für $\pi_k \in P_m(\varepsilon)$

$$\frac{|\tilde{\pi} - \pi_{k+1}|}{|\tilde{\pi} - \pi_k|} \leq 1 - \frac{1}{m} \cdot \delta_\varepsilon,$$

also gilt

$$|\tilde{\pi} - \pi_k| \leq 2(1 - \frac{\delta_\varepsilon}{m})^k \quad (|\tilde{\pi} - \pi| \leq 2),$$

so lange $\pi_1, \pi_2, \pi_3, \dots, \pi_k \in P_m(\varepsilon)$.

Wegen $\delta_\varepsilon > 0$ ist es immer möglich, k_ε so zu wählen, daß

$$2 \cdot (1 - \frac{\delta \varepsilon}{m})^k \varepsilon < \varepsilon.$$

Also muß für beliebiges π wenigstens einer der Punkte

$$\pi, \pi_1, \pi_2, \pi_3, \dots, \pi_{k_\varepsilon}$$

außerhalb von $P_m(\varepsilon)$ liegen. Damit ist aber Behauptung (2) bewiesen.

Aus Behauptung (1) und (2) folgt, daß für beliebiges π die Ungleichung

$$|\pi \tilde{\pi} - \pi \cdot \pi_{k_\varepsilon}| < 2\varepsilon$$

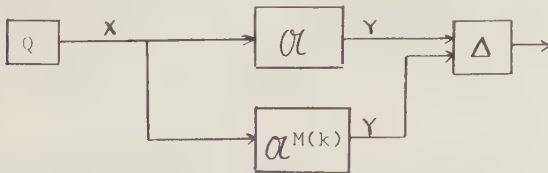
gilt, da $|\pi \tilde{\pi} - \pi \cdot \pi_k| \cong |\pi \tilde{\pi} - \pi \pi_{k+1}|$ (Monotonie, Hilfssatz 4.4),

womit der Satz bewiesen ist.

Es sei hier noch angemerkt, daß die π_k keineswegs gleichmäßig gegen $\tilde{\pi}$ konvergieren.

4.3 Vorhersage eines Automaten mit Multiplex-Automaten

Wir betrachten jetzt die Vorhersageanordnung:



Wenn $\mathcal{A}$ nach Eingabe von $\bar{x} \in X^*$ in der Ausgabeverteilung $\pi_{O A \bar{x}}^B$ weilt, so weilt $\mathcal{A}^{M(k)}$ in der Ausgabeverteilung

$$(\pi_{O A \bar{x}}^B)_k,$$

wobei k die k -Iteration symbolisiert. Wir haben also mit $w =$ Übertragungsfunktion von $\mathcal{A}$, $w_k =$ Übertragungsfunktion von $\mathcal{A}^{M(k)}$ (vgl. Satz 3.5 und Beweis dazu)

$$f(\mathcal{A}, \mathcal{A}^{M(k)}, Q, t) = \sum_{\bar{x} \in X^t} w_Q(\bar{x}) \sum_{\substack{\bar{y} y \in Y^t \\ \bar{y}' y' \in Y^t}} w(\bar{y} y | \bar{x}) w_k(\bar{y}' y' | \bar{x}) \cdot \Delta(y, y'),$$

und das läßt sich vereinfachen auf

$$f(\mathcal{A}, v_{\mathcal{A}^{M(k)}, Q, t}) = \sum_{\bar{x} \in X^t} w_{\mathcal{A}}(\bar{x}) \sum_{y, y' \in Y} v(y|\bar{x}) v_k(y'|\bar{x}) \Delta(y, y'),$$

wobei $v(y|\bar{x})$ wie im Beweis zu Satz 3.5 für $\mathcal{A}$, $v_k(y'|\bar{x})$ analog für $\mathcal{A}^{M(k)}$.

Wenn $\pi = \pi_{\mathcal{O} \bar{x}}$ die Ausgabeverteilung von $\mathcal{O}$ nach der Eingabe von $\bar{x}$ ist, so sind die Komponenten von π gerade die Wahrscheinlichkeiten $v(y|\bar{x})$ für $y = y_1, y_2, \dots, y_m$. Analog gibt π_k die Ausgabewahrscheinlichkeiten $v_k(y|\bar{x})$ von $\mathcal{A}^{M(k)}$ an. Es gilt also, wie man sich leicht überlegt,

$$\begin{aligned} \sum_{y, y' \in Y} v(y|\bar{x}) \cdot v_k(y'|\bar{x}) \cdot \Delta(y, y') &= 1 - \sum_{y \in Y} v(y|\bar{x}) \cdot v_k(y|\bar{x}) \\ &= 1 - \pi \cdot \pi_k. \end{aligned}$$

Wir erhalten also

$$\frac{f(\mathcal{A}, v_{\mathcal{A}^{M(k)}, Q, T})}{T} = \sum_{(\pi', \pi) \in D(T)} |\pi| \cdot \left(1 - \frac{\pi_B}{|\pi_B|} \cdot \left(\frac{\pi_B}{|\pi_B|} \right)_k \right),$$

wobei $D(T)$ wie im Beweis zu Satz 3.7 aus Q und $\mathcal{A}$ gewonnen wird. Dabei ist

$$\frac{\pi_B}{|\pi_B|}$$

eine (auf die Länge 1 normierte) Ausgabeverteilung von $\mathcal{O}$,

$$\left(\frac{\pi_B}{|\pi_B|} \right)_k$$

die k -Iteration dieser Ausgabeverteilung. Wenn für beliebige positive Vektoren π die k -Iteration π_k definiert wird gemäß

$$\pi_k := \left(\frac{\pi}{|\pi|} \right)_k \quad (|\pi| > 0),$$

so können wir schreiben

$$\frac{f(\mathcal{A}, v_{\mathcal{A}^{M(k)}, Q, T})}{T} = \sum_{(\pi', \pi) \in D(T)} (|\pi| - (\pi_B) \cdot (\pi_B)_k),$$

oder, wenn $D(T)OB$ die Dichte

$$D(T)OB := \{ \pi_B \mid (\pi', \pi) \in D(T) \}$$

bezeichnet (B = Markierungsmatrix von $\mathcal{A}$),

$$\frac{F(\mathcal{A}, v_{\mathcal{A}^{M(k)}, Q, T})}{T} = \sum_{\pi \in D(T)OB} |\pi| - \pi \cdot \pi_k.$$

Satz 4.2

Sei Q eine Quelle und $\mathcal{A}$ ein Automat. Dann gilt

$$\lim_{k \rightarrow \infty} \tilde{F}(\mathcal{A}, v_{\mathcal{A}^{M(k)}, Q}) = \tilde{F}_{\min}(\mathcal{A}, Q).$$

$$\begin{aligned} \text{Beweis: Wegen } \lim_{k \rightarrow \infty} (|\pi| - \pi \cdot \pi_k) &= \lim_{k \rightarrow \infty} |\pi| \left(1 - \frac{\pi}{|\pi|} \cdot \pi_k\right) = \\ &= |\pi| \cdot \left(1 - \text{MaxKomp}\left(\frac{\pi}{|\pi|}\right)\right) = |\pi| - \text{MaxKomp}(\pi) \end{aligned}$$

(Hilfssatz 4.4) haben wir

$$\begin{aligned} \lim_{k \rightarrow \infty} \frac{F(\mathcal{A}, v_{\mathcal{A}^{M(k)}, Q, T})}{T} &= \lim_{k \rightarrow \infty} \sum_{\pi \in D(T)OB} \frac{|\pi| - \pi \cdot \pi_k}{T} = \\ &= \sum_{\pi \in D(T)OB} \frac{|\pi| - \text{MaxKomp}(\pi)}{T} = F(\mathcal{A}, v_{\text{opt}, Q, T})/T. \end{aligned}$$

Da wir in Satz 4.1 die gleichmäßige Konvergenz von $|\pi \tilde{\pi} - \pi \cdot \pi_k|$ (bezogen auf alle π) bewiesen haben, konvergiert auch

$$\begin{aligned} \lim_{k \rightarrow \infty} \left[\frac{F(\mathcal{A}, v_{\mathcal{A}^{M(k)}, Q, T})}{T} - \frac{F(\mathcal{A}, v_{\text{opt}, Q, T})}{T} \right] &= \\ \lim_{k \rightarrow \infty} \sum_{\pi \in D(T)OB} |\pi| - \pi \cdot \pi_k - |\pi| + \text{MaxKomp}(\pi) &= \\ = \lim_{k \rightarrow \infty} \sum_{\pi \in D(T)OB} \text{MaxKomp}(\pi) - \pi \cdot \pi_k &= 0 \end{aligned}$$

gleichmäßig für alle möglichen $D(T)OB$.

Also gilt, da die Limites für $T \rightarrow \infty$ und $k \rightarrow \infty$ wegen der gleich-

mäßigen Konvergenz vertauscht werden dürfen:

$$\begin{aligned}
 \lim_{k \rightarrow \infty} \tilde{F}(\mathcal{A}, v_{\mathcal{A}^{M(k)}, Q}) &= \lim_{k \rightarrow \infty} \lim_{T \rightarrow \infty} \sum_{\pi \in D(T) \cap \mathcal{O}B} |\pi| - \pi \cdot \pi_k = \\
 &= \lim_{T \rightarrow \infty} \lim_{k \rightarrow \infty} \sum_{\pi \in D(T) \cap \mathcal{O}B} |\pi| - \pi \cdot \pi_k = \\
 &= \lim_{T \rightarrow \infty} \sum_{\pi \in D(T) \cap \mathcal{O}B} |\pi| - \text{MaxKomp}(\pi) = \tilde{F}(\mathcal{A}, v_{\text{opt}}, Q) = \tilde{F}_{\min}(\mathcal{A}, Q).
 \end{aligned}$$

Damit ist Satz 4.2 bewiesen.

Korollar 4.1

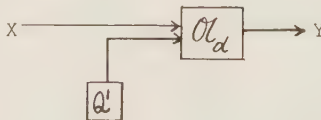
Sei $\mathcal{A}$ ein Automat, Q eine speisende Quelle. Zu jedem $\varepsilon > 0$ existiert ein (endlicher) Automat $\mathcal{A}'$, so daß

$$\tilde{F}(\mathcal{A}, v_{\mathcal{A}', Q}) - \tilde{F}_{\min}(\mathcal{A}, Q) < \varepsilon$$

Es gilt also auch $\tilde{F}_{\min}(\mathcal{A}, Q) = \tilde{F}_{\text{stoch}}(\mathcal{A}, Q)$.

Die Herleitung dieses Resultates liefert gleichzeitig eine Technik, mit der man für einen beliebigen Automaten $\mathcal{A}$ einen "relativ guten" Vorhersageautomaten konstruieren kann. Dies geschieht oben durch die Multiplex-Technik, genauer durch Automaten des Typs $\mathcal{A}^{M(k)}$.

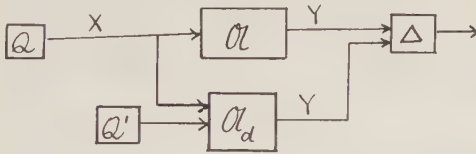
Die Anwendung eines Satzes von NIEH und CARLYLE [5] zusammen mit Korollar 4.1 liefert ein weiteres Resultat, das kurz erläutert werden soll. Der Satz von NIEH und CARLYLE besagt, daß jede Übertragungsfunktion $w(\cdot|\cdot)$, die in einem endlichen stochastischen Automaten realisiert werden kann, auch von einem System der Gestalt



erzeugt werden kann, wobei $\mathcal{A}_d$ ein deterministischer Automat, allerdings mit nichtdeterministischer initialer Zustandsverteilung. Die Quelle Q' ist gedächtnisfrei, d. h. jedes ausgegebene Zeichen

ist vollständig unabhängig von vorausgegangenen oder zukünftigen Signalen (Zufallslieferant).

Korollar 4.1 liefert somit die Aussage, daß mit einer Vorhersageanordnung der Gestalt



mit deterministischem O_d und gedächtnisfreiem Q' der Vorhersagefehler beliebig nahe an $\tilde{F}_{\min}(O, Q)$ heruntergedrückt werden kann. In systemtheoretischer Sprechweise würde man O_d als deterministisches Filter mit Gedächtnis (Zustände!) und Q' als Rauschgenerator bezeichnen, während O ein nichtdeterministisches Filter mit Gedächtnis und Δ ein Differenzenglied wäre.

L i t e r a t u r

- [1] Booth, T. L.
Estimation, Prediction and Smoothing in Discrete Parameter Systems.
Report GD/EB-22, University of Connecticut, School of Engineering (1967).
- [2] Doob, J. L.
Stochastic Processes.
J. Wiley, New York 1953.
- [3] Kemeny, J. G., Snell, J. L.
Finite Markov Chains.
Van Nostrand, Princeton 1960.
- [4] Ljusternik, L. A., Sobolew, W. I.
Elemente der Funktionalanalysis.
Akademie-Verlag, Berlin 1968.
- [5] Nieh, T. T., Carlyle, J. W.
On the Deterministic Realization of Stochastic Finite-State Machines.
Proceedings of the "Second Annual Princeton Conference on Information Sciences and Systems" 1968, 189-193.
- [6] Starke, H. P.
Theorie stochastischer Automaten.
Elektr. Informationsverarbeitung und Kybernetik 1(1965), 5-32, 71-98.
- [7] von Neumann, J.
Probabilistic Logics and the Synthesis of Reliable Organisms From Unreliable Components.
Automata Studies, University Press, Princeton 1956.
- [8] Wolf, F.
Untersuchungen über den Vorhersagefehler bei stochastischen Automaten.
Dissertation, Hannover 1968.

Begriffsregister (mit Seitenangabe)

Absolutbetrags-Metrik 17
Alphabet 8
Äquivalenz von Automaten 11
Äquivalenz von Dichten 31
Äquivalenzklassen von Dichten 31
Ausgabealphabet 9
Automat, initialer stochastischer Moore-Automat mit
deterministischer Markierung 9
bedingte Überführungsmatrizen einer Quelle 37
bedingte diskrete Wahrscheinlichkeitsfunktion 9
bedingte Wahrscheinlichkeit 9
deterministisches Filter mit Gedächtnis 65
Dichte
 Mengenschreibweise für 38
 Definition 31
 diskrete 35
 kontinuierliche 35
Differenzenglied 65
diskretes Wahrscheinlichkeitsmaß 8
diskrete Zeit 10
distributiv 41
Eingabealphabet 9
Ergodizität 7
gedächtnisfrei 64
Häufungspunkt D 44
Hauptvektoren 32
Hausdorff, Satz von 35
k-Iteration von Vektoren 54
k-Multiplex-Automat 53
kompakt 35
Kurzzeitverhalten einer Vorhersagestrategie 49
Markierungsmatrix 16
Markow-Ketten 15
Masse einer Dichte 31
massentreue Operatoren 41
Maßzahl für Determiniertheit 5
Metrik $\triangle$ 19
metrischer Raum 17

Multiplex-Automat 50
Multiplex-Technik 6
nichtdeterministische Markierung 11
Operator, Matrix als 17
 schwach kontrahierend 18
Parallelschaltung 13
Q- γ -Dichte 37
Quelle 14
Quellenwahrscheinlichkeit 14
Quellenalphabet 15
Raum der Dichten 30
Rauschquelle 6
Repräsentant, von Dichten 31
Sampling 7
sequentielle Teilung der Wahrscheinlichkeiten 20
Serienprodukt 12
stochastischer Automat 9
stochastische n-n-Matrix 15
Taktoperator A 40
Übertragungsfunktion 11
Verhalten 11
vollständiger Raum der Dichten 35
Vorhersageanordnung 19
Vorhersagefehler, mittlerer 4, 21
Vorhersagestrategie 3,20
 optimale 5, 28
 mit endl. Automaten
 realisierbare 22
Zustandsverteilung, initiale 9

Wichtige Formelzeichen

(in der Reihenfolge, in der sie im Text auftreten)

<u>Symbol</u>	<u>Seite</u>	<u>Symbol</u>	<u>Seite</u>
X	8	$V_{\mathcal{A}}(\bar{y} \bar{x})$	22
$\bar{x}$	8	$\tilde{L}(x_j)$	23
$\ell(\bar{x})$	8	$\tilde{F}_{\text{stoch}}(\mathcal{A}, Q)$	25
Λ	8	$\tilde{F}_{\text{min}}(\mathcal{A}, Q)$	25
X^k	8	$v(j \bar{x})$	27
$P(M)$	9	$Y_{\bar{x}}$	27
$\mathcal{A}$	9	V_{opt}	28
Y	9	D	31
Z	9	$\mathcal{S}$	32
w	9	$A_Q(x)$	37
μ	9	D_t	37
π_o	9	$\text{MaxKomp}(\pi)$	38
$w(z' z, x)$	10	$f(\mathcal{A}, V, Q, t)$	38
$w(\bar{y} \pi, x)$	11	$A(\text{Taktop.})$	40
$w(\bar{y} \bar{x})$	11	$D(T)$	44
$\mathcal{A}_1 \rightarrow \mathcal{A}_2$	12	$\mathcal{A}^M$	50
$\mathcal{A}_1 \parallel \mathcal{A}_2$	13	$\mathcal{A}^{M(k)}$	53
Q	14	π_k	54
w_Q	14	$\tilde{\pi}$	54
A_x	15	η	54
$A_{\bar{x}}$	16	δ	55, 56
B	16	$\pi_1 \cdot \pi_2$	57
B_j	16	P_m	
v_j	16	$P_m(\mathcal{E})$	60
$\pi_o A_{\bar{x}} B_j$	17	$\delta_{\mathcal{E}}$	60
$P(Z)$	17	$D(T)OB$	63
$P(n)$	17		
$ \pi - \pi' $	17		
Δ	19		
$V(\bar{y} \bar{x})$	20		
$F(\mathcal{A}, V, Q, T)$	21		
$\tilde{F}(\mathcal{A}, V, Q)$	21		

Bisher sind folgende Arbeitsberichte des Instituts für
Mathematische Maschinen und Datenverarbeitung erschienen:

BAND 1

- Nr. 1 SCHMITT, A.:
Theorie der nichtdeterministischen und unvollständigen
Mealy-Automaten. (Januar 1967)
- Nr. 2 GÖHRING, W., HENNING, W., RASCHE, W.:
Untersuchungen zur Kommunikation und Mustererkennung,
Teil 1: Übersicht. (Juli 1967)
- Nr. 3 NAMNECK, P.:
Untersuchungen zur Kommunikation und Mustererkennung,
Teil 2: Über Verfahren zur syntaktischen Analyse von
Sätzen formaler Sprachen. (Juli 1967)
- Nr. 4 EGGERS, B.:
Untersuchungen zur Kommunikation und Mustererkennung,
Teil 3: Möglichkeiten der automatischen Erkennung und
selbständigen Korrektur von syntaktischen Fehlern
in formalen Sprachen. (Januar 1968)
- Nr. 5 SCHNEIDER, H. J.:
Programmierung numerischer Probleme. (Oktober 1967)
- Nr. 6 WOLF, F.:
Untersuchungen über den Vorhersagefehler bei stochasti-
schen Automaten. (Februar 1968)
- Nr. 7 KLAR, R.:
Rechenautomaten I. (April 1968)
- Nr. 8 SCHMITT, A., FEHLING, P.:
Rechenautomaten II. (November 1968)
- Nr. 9 WEISS, S.:
Untersuchungen zur Kommunikation und Mustererkennung,
Teil 4, Ein Programmsystem zum Testen von Mustererken-
nungsverfahren unter Verwendung eines Sichtgerätes.
(Juli 1968)
- Nr. 10 Dokumentation über die Arbeit des Instituts in den
Jahren 1966 bis 1968. (Januar 1969)

BAND 2

- Nr. 1 VOLLMAR, R.:
Über einen Automaten mit Speicherband.
(Januar 1969)
- Nr. 2 HÄNDLER, W.:
Design Philosophy for Computer Systems.
(März 1969)
- Nr. 3 BRACKMANN, U., HÄNDLER, W., LANGE, M., RASCHE, W.,
WOLF, H. P.:
Beiträge zur Kommunikation Mensch-Maschine.
(April 1969)
- Nr. 4 HAFERKORN, J., SCHMITT, A., SEIDEL, R., WOLF, F.:
Rechnergestützter Unterricht. (April 1969)
- Nr. 5 BÖHME, J. F., GÜHRING, W.:
Zwei Beiträge zur Mustererkennung. (August 1969)
- Nr. 6 KOEHL, J.:
Datenreduktion zur Speicherung ebener Figuren.
(April 1969)
- Nr. 7 HAFERKORN, J., HENNING, W., RASCHE, W.:
Ein bildschirmorientiertes Betriebssystem für die
PDP 7.
- Nr. 8 SAFFERT, J.:
Mikroprogramm-Steuerwerke. (Juni 1969)
- Nr. 9 SPIES, P. P.:
Rechnerstrukturen und Ablaufgeschehen im Rechen-
automaten - Analysis und Simulation. (September 1969)
- Nr. 10 HÄNDLER, W., WOLF, F.:
Bericht über eine USA-Reise - Aktuelles aus dem
Gebiet der Informatik. (September 1969)
- Nr. 11 Dokumentation über die Arbeit des Instituts im
Jahre 1969. (Januar 1970)

BAND 3

- Nr. 1 SCHMITT, A.:
Bildschirme und Kommunikation Mensch-Maschine
(Februar 1970)
- Nr. 2 BÖHME, J.F.:
Experimente mit stochastischen Automaten
(Juli 1970)
- Nr. 3 SCHNEIDER, H.J.:
Chomsky-Systeme für partielle Ordnungen
(August 1970)
- Nr. 4 HAAS, G.:
Programme zur dialogischen Logik
(Oktober 1970)
- Nr. 5 WEICKER, Reinhold:
Erkennung von kontextfreien Sprachen
durch Tabulator- Turingmaschinen
(Juli 1970)

Offsetdruck-Fotodruck

J. Hogl, 8520 Erlangen Hauptstr. 109, (Nähe Martin-Luther-Platz) Tel. 22990
